The Secondary Structure of The 5'-end of The STb mRNA Affects Its Secretion Efficiency

By

Eyad Kinkar

A thesis submitted in partial fulfillment of the requirements

for the degree of Master of Science (MSc) in Biology

The Faculty of Graduate Studies

Laurentian University

Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis
Titre de la thèse: The Secondary Structure of The 5'-end of The STb mRNA Affects Its Secretion Efficiency

Name of Candidate
Nom du candidat: Kinkar, Eyad

Degree
Diplôme: Master of Science

Department/Program
Département/Programme: Biology

Date of Defence
Date de la soutenance December 15, 2016

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Mazen Saleh
(Supervisor/Directeur(trice) de thèse)

Dr. Rob Lafrenie
(Committee member/Membre du comité)

Dr. Jeffrey Gagnon
(Committee member/Membre du comité)

Dr. Mark Bayfield
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. Shelley Watson
Madame Shelley Watson
Acting Dean, Faculty of Graduate Studies
Doyenne intérimaire, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

ABSTRACT

The majority of secreted proteins in *E. coli* are targeted to and translocated across the cytoplasmic membrane through a signal sequence within the N-terminus of the protein. The major membrane complex facilitating this process is the general secretory or Sec-dependent membrane complex, also known as T2SS secretion. Research with proteins that are processed for export through a different pathway, called the T3SS secretion system, has recently challenged this widely accepted hypothesis. This research has provided evidence that the T3SS secretion system substrates may use mRNA sequences as a secondary means of protein targeting in addition to the N-terminus of the synthesized protein. The main objective of this study was to test whether or not the mRNA of the T2SS secretion system substrates may have structural information sufficient for efficient targeting of the protein for secretion. A well-studied substrate of the T2SS secretion system in *E. coli* is the heat-stable enterotoxin B (STb). The STb gene was cloned and a number of mutations were introduced within the 5' end of its mRNA to look at the effect of such mutations on protein secretion. The results showed that mutations that lead to changes in the AT/GC ratio within this 5’ region of the gene, while conserving the amino acid sequence of that region, affected toxin secretion. Bioinformatic analysis showed that these mutations affected the secondary structure of the 5' end of the mRNA and indicated that significant changes can be produced in the secondary structure of the mRNA when the AT/GC ratio is modified. This data shows that alterations in the secondary structure of the 5' end of the mRNA, without changes in the amino acid sequence of the N-terminus of the protein, can affect the efficiency with which the toxin is secreted and hints to a possible role for the mRNA in targeting proteins for export across the cytoplasmic membrane.

ACKNOWLEDGMENTS

Thank God for all the blessings.

There are no words can describe my grateful gratitude and feelings to all those who contributed and supported me in different ways throughout my studies. First, it is my pleasure to express my warm appreciations to my supervisor, Dr. Mazen Saleh, who believed in me and in the success of the present study. His diverse research lab, knowledge, and most importantly expertises added considerably more flavor to my scientific experience. My warm appreciations are also extended to my committee members, Dr. Robert Lafrenie and Dr. Jeffrey Gagnon who devoted their valuable time and guidance at all levels throughout my graduate studies. I must also acknowledge Dr. Paul Michael and My sister Ayat Kinkar for their theoretical innovations and supports during my research.

It would be impossible to forget extending my gratitude to my family for their undeniable love and patience during my ups and downs and their unstoppable support and encouragements through my entire life. I also would like to send special gratefulness to Dr. Pejman Hanifi-mogadum for being a great mentor in research. I must also genuinely acknowledge Dr. Assam Kumar and Dr. Kabwe Nkongolo as their labs were being a magnificent help throughout my scientific journey. I would also like to express my best wishes and thanks to all of my friends who I met here and being a family to me.

Last but not least, a massive recognition to My Father Abdulrahman Kinkar God bless his soul who made my education in Canada is financially possible as well as the Saudi Cultural Bureau.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

bp	Base pairs
kDa	Kilo Dalton
MW	Molecular weight
RT	Room temperature
rpm	Rotations per minute
g	Gravity
U	Unit
Abs	Absorbance
C	Carboxyl-terminal of protein
N	Amino-terminal of protein
LB	Luria Bertani
AmBic	Ammonium bicarbonate
EDTA	Ethylenediaminetetracetic acid
PAGE	polyacrylamide gel electrophoresis
TEMED	Tetramethylethylenediamine
TAE	Tris-acetate-EDTA buffer
Tricine-SDS-PAGE	Tricine Sodium dodecyl sulfate polyacrylamide gel electrophoresis
E. coli	*Escherichia coli*
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	messenger RNA
IM	Inner membrane
OM	Outer membrane
MP	Membrane protein
Sec	Secretory
Tat	Twin-Arginine translocation

RBP	RNA binding protein
NCBI	National Center for Biotechnology Information
PCR	Polymerase chain reaction
Ta	Annealing temperature
Tm	Melting temperature

INTRODUCTION

Bacterial Cell Wall

Most bacterial cells present a distinctive cell wall architecture that provides mechanical support to the cell and contributes to maintaining its shape, (Huang et al., 2008; Osborn, 1969). The cell wall possesses a complex polymer consisting of acetamido sugars and a few distinct amino acids. There are many names for this polymer including glycopeptide, glycosaminopeptide, mucopeptide, mucopolymer, murein, and peptidoglycan. However, peptidoglycan is widely used because it is believed to be the best term to describe the bacterial sacculus structurally (Harz, et al., 1990). Although the sacculus is well known to completely shield the bacterial cell from osmotic and mechanical lysis, as well as filtering nutrients and toxins passing though the cell membrane, microbiologists are still questioning the specific architecture of the cell wall (Gan, et al., 2008).

The cell wall in Gram-negative microorganisms such as *Escherichia coli* is composed of three layers. It contains a thin peptidoglycan layer sandwiched between an outer membrane and a cytoplasmic membrane (Figure 1) (Beveridge, 1999; Kellenberger & Ryter, 1956). The periplasmic space between the two membranes contains the periplasm, a matrix that is distinct in composition from the cytoplasm and the outer milieu (Kellenberger & Ryter, 1956; Osborn, 1969). The periplasmic space is actually essential within the cell wall of the Gram-negative bacteria because it contains a reduced environment for the proper folding of exported proteins. Collectively, the extracellular membrane, periplasm, peptidoglycan layer and plasma membrane comprise the Gram-negative sacculus (Beveridge, 1999; Kellenberger & Ryter, 1956).

The Gram-negative cell wall is unlike the Gram-positive cell wall in that the latter does not have an outer membrane and the peptidoglycan layer is thicker than that observed in the Gram-

negative organisms. According to recent electron cryotomographic analysis, the average profile measurement of two Gram-negative bacterial cell walls (*E. coli* and *Caulobacter crescentus*) is about ~4 nm (Harz et al., 1990; Yao, et al., 1999). The cell wall in Gram-negative bacteria does not share the exact thickness in all genera. Research has reported that the cell wall of *E. coli* is greater than 9-nm to accommodate the average length of the glycan strands and the peptidoglycan layer whereas *C. crescentus* reveals a different density (~7 nm). These variations are likely due to variations in the abundance of proteins embedded in the sacculus (Gan et al., 2008; Harz et al., 1990; Yao et al., 1999). Despite the differences in thickness, all Gram-negative cell walls are considered to be a barrier to regulate passage of ions, energy-consuming processes, proteins and other macromolecules that are synthesized in the cytosol and fulfill their extended metabolic activities or structural functionalities within or outside the cell. Different transport mechanisms are utilized to allow the passage of different molecules through the cell wall (Schnaitman, 1970). The passage of secretory proteins through the different cell wall barriers is performed without compromising protein structure and function and is controlled by various transport systems that are involved in the recognition and localization of the targeted proteins (Costa et al., 2015; Schnaitman, 1970).

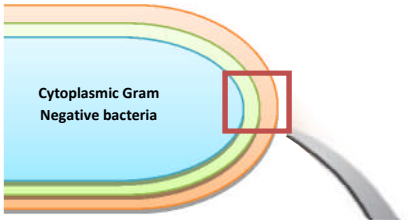

Figure 1: The cell-wall structure of Gram Negative bacteria that show the three different main barriers including outer membrane, cytoplasmic membrane, and the periplasmic space. Adopted from http://faculty.ccbcmd.edu/~gkaiser/SoftChalk%20BIOL%20230/Prokaryotic%20Cell%20Anatomy/Gram-negative%20cell%20wall/Gram_negative_cell_wall/Gram_negative_cell_wall_print.html

Protein Synthesis and Targeting

This section will present an overview of protein synthesis in Gram-negative bacteria to facilitate the description of the mechanisms involved in protein targeting to the membrane for subsequent translocation. Protein synthesis, or translation, is defined as the process by which mRNA is translated into a protein with the aid of the ribosome (Pelletier & Sonenberg, 1987; Thanaraj & Pandit, 1989). In 1953, the DNA double helix structure was described and confirmed that it had a distinctive identity from the single-stranded RNA molecule (Franklin & Gosling, 1953; Pelletier & Sonenberg, 1987).

Although the processes of transcription and translation that result in protein synthesis can take place simultaneously in bacteria, the processes of translation can be divided into three stages: initiation, elongation, and termination. The first step of translation initiation in *E. coli* involves the synthesis of various components to assemble the required complex for translation, these include IF1, IF2, and IF3 which bind to the 30S ribosomal subunit, tRNA, and mRNA (de Smit & van Duin, 1990; Marintchev & Wagner, 2004). Translation starts when the 16S Ribosomal RNA (rRNA), a part of the 30S ribosomal subunit, encounters the Shine-Dalgarno (SD) sequence in the bacterial mRNA. This leads to the dissociation of IF2 from the 30S ribosome to promote binding of the larger ribosomal subunit 50S to allow assembly of the a complete functional ribosomal complex (de Smit & van Duin, 1990; Kaczanowska & Rydén-Aulin, 2007; Pape, et al., 1998). The ribosome assembles on the mRNA and the tRNA-fmet, which is linked to a formylated methionine residue, interacts with the P-site of the ribosome complex to initiate the beginning of protein synthesis. The next stage is protein elongation. This involves the incorporation of amino acids linked to aminoacylated tRNA into the A-site of the ribosome complex and peptide bond formation between adjacent amino acids. The growing peptide is then transferred to the P-site of the ribosome complex and the next amino acyl-tRNA is incorporated into the A-site to repeat the process. The

third and final stage is the termination of synthesis. This occurs when the releasing factors (RF1 and RF2) are incorporated into the ribosomal complex because they bind to the termination codons. This results in the termination of protein synthesis and the disruption of the ribosomal complex and release of the 30S and 50S subunits in the cytoplasm. A single RNA can be translated many times to create protein. However, after recognition of some signal the mRNA can be degraded so that the nucleotides can be recycled (de Smit & van Duin, 1990; Malys & McCarthy, 2011; Marintchev & Wagner, 2004; Pape et al., 1998). All proteins (secreted or cytoplasmic) in *E. coli* are produced in the cytoplasm and depending on the specific export signal, can be exported. The exported substrates have three possible target localizations: the protein could remain anchored within the bacterial outer or inner membrane, it can remain in the periplasm, or it can be released into the extracellular milieu as a secreted protein (Gerlach & Hensel, 2007). The targeting system employs a few distinct pathways that can be utilized by *E. coli* to translocate all secretory proteins. Some secretory proteins are also known as effector proteins because they have the ability to affect the host cell and allow the adherence of virulence factor to host cells in specialized secretion systems of virulent bacteria. They can also help in the colonization and multiplication of pathogens in the host cells (Gerlach & Hensel, 2007).

In *E. coli*, up to 40% of all cellular proteins are either; exported to the periplasm, are present in the OM, or are freely secreted through a number of secretion systems. The T2SS (general secretory system) is the best described export system and involves recognition for the N-terminal sequence of the secreted protein to direct export. However, since the late 1980s, it became clear that the T2SS secretion system is not the only translocase present in the cytoplasmic membrane (Costa et al., 2015; Natale, et al., 2008). *E. coli* has a very sophisticated one-step or two-step secretion mechanism to translocate various virulence factors across their cell envelope into the

extracellular environment or into the host cells. These secreted virulence factors include small molecules, proteins, and DNA (Rêgo, et al. , 2010). Seven different secretion systems have been discovered thus far (T1SS through T7SS). Research separates them into two major categories; the one-step secretion systems and the two-step secretion system. T1SS, T3SS, T4SS, T5SS, and T6SS are one-step secretion systems that translocate polypeptides directly from the cytoplasm into the extracellular space (Costa et al., 2015; Gerlach & Hensel, 2007). However, T2SS utilizes a two-step secretion system where polypeptides get translocated through the inner membrane into the periplasmic space and then subsequently get secreted into the extracellular space through the outer membrane (Costa et al., 2015; Gerlach & Hensel, 2007; Natale et al., 2008). Finally, T7SS is restricted to Mycobacteria and has not been reported in any other bacteria thus far. Most of the one-step secretion systems translocate partially folded or cytoplasmically folded proteins, but the two-step secretion system secretes mostly unfolded proteins (Costa et al., 2015; Gerlach & Hensel, 2007).

T2SS secretion in *E. coli*

There are two distinct pathways utilized by T2SS systems to export proteins across the membrane of *E. coli*: known as the Sec-dependent and Sec-independent pathways. The Sec-dependent pathway system utilizes a two-step secretion mechanism, whereby unfolded substrates are first translocated into the periplasmic space by IM-spanning transporters, and are then transferred to the OM or secreted into the extracellular space by a dedicated OM-spanning secretion system (Lycklama et al., 2012; Nivaskumar & Francetic, 2014; Palmer & Berks, 2012). In this system, research has shown that the secretory proteins are exported across the membrane by a channel complex called the heterotrimeric SecYEG protein complex. This process is directed by the Signal Recognition Particle (SRP) (a ribonucleoprotein complex) or by SecB (a chaperone

that is specific for secretion) while maintaining the proteins in their unfolded state (Figure 2) (Brundage, et al., 1990; Driessen, et al., 2001). SecYEG is a complex of three proteins (SecY, SecE, and SecG) that is the central component of the translocase. For a successful polypeptide translocation, the process also requires association with the cytosolic ATPase, SecA, the ribosome, YidC and SecDFyajC, another heterotrimeric membrane protein complex (Duong & Wickner, 1997). Since secretory proteins are synthesized in the cytoplasm and are then transported into the periplasmic space where they can be properly folded before being secreted into the extracellular environment, the system requires between twelve and sixteen different proteins and enzymes (Brundage et al., 1990; Costa et al., 2015; Driessen et al., 2001; Natale et al., 2008).

An alternative system for transportation of proteins has also been identified in *E. coli*. This system is known as the Tat system (*t*win-*a*rginine *t*ranslocase) or the Sec-independent pathway. This system is more restricted in scope and nature than the Sec-dependant pathway because it transports only certain folded proteins (those containing cofactors) across the membrane to be secreted or embedded in the IM (Natale et al., 2008; Palmer & Berks, 2012). This pathway is most prominent in specific strains of *E. coli*, where protein folding is incompatible with the Sec-dependent pathway (Figure 2). In *E. coli* almost 6% of all secreted proteins are translocated through this pathway. The redox enzymes are typical substrates for the Tat pathway and they require cofactors to assemble into a complete functional complex (Ize et al., 2002; Sargent, et al., 1999). The integral components of this system are TatA, TatB, and TatC. The Tat system uses its nano-machinery and the energy of proton motive force to translocate proteins across the cytoplasmic membrane (Greene et al., 2007; Yen, et al.,, 2002).

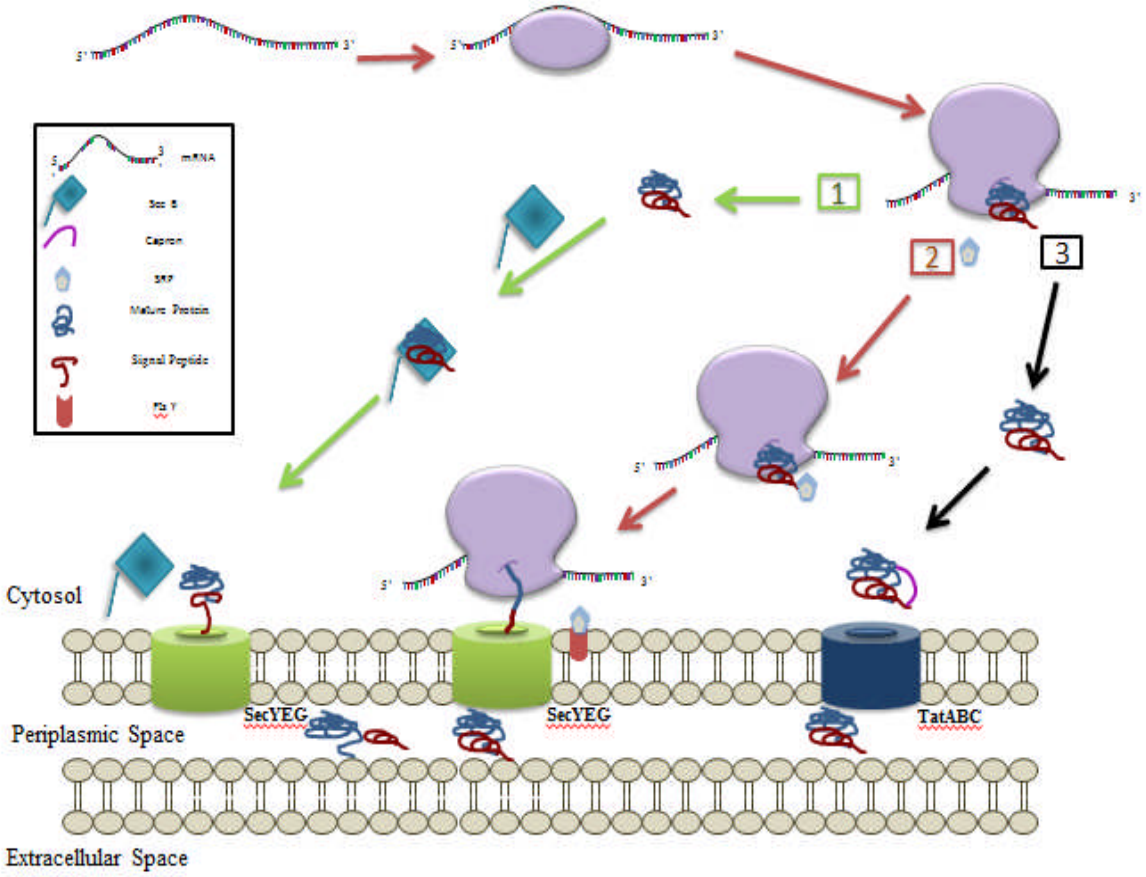

Figure 2: The three different and most recognized pathways of the T2SS secretion system in Gram-negative bacteria. The translocation of proteins through the T2SS secretion system is mainly controlled by one of three different pathways: 1- SecB-dependent pathway, 2- signal recognition particle (SRP) pathway 3- twin-arginine translocation (TAT) system. Adopted from http://www.athenaes.com/tech_brief_ACESyebf.php

The proper sorting and targeting of proteins to their final cellular destination, requires the presence of a signal sequence. In *E. coli* most secretory proteins are engaged with a cleavable signal called the N-terminal signal sequence that guides them to the periplasmic space following translocation across the membrane. The signal peptide is usually cleaved off from the rest of the protein during the translocation step by the signal peptidase. The mature protein is subsequently folded and secreted into the extracellular environment.

The N-terminal Signal Sequence

Exported proteins are usually anchored within the inner membrane, the outer membrane, the periplasm, or transferred to the extracellular environment of the cell through the secretion process. Studies have shown that secretion systems require a signal peptide for protein secretion is present on the N- terminal in both eukaryotes and prokaryotes. The signal peptide is normally between 5 and 30 amino acids long within the N-terminal region of the protein (Luirink & Sinning, 2004; Papanikou, et al., 2007).

The signal peptide is divided into three regions. The first region is a short positively charged region that is close to the N-terminus of the protein and it is called the n-region. It is believed that the n-region could possibly be involved in enforcing a proper structure on the signal peptide during translocation. The second, medial stretch of amino acids, is about 16 amino acids in length and represents the core hydrophobic region (h-region). Lastly, the c-region that contains the signal cleavage site (Figure 3) (Luirink & Sinning, 2004; Papanikou et al., 2007; Randall & Hardy, 2002; Ullers et al., 2004). The cleavage site is recognized by the signal peptidase that is responsible for

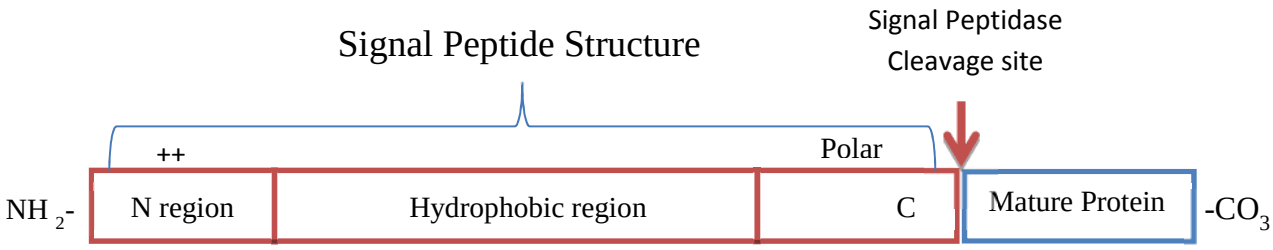

Figure 3: Characterization of the N-terminal signal peptide of *E. coli's* secreted proteins that utilise T2SS secretion system.

cleaving the signal peptide after completion of translocation to generate the mature protein (Luirink & Sinning, 2004; Papanikou et al., 2007). All three regions are required for efficient sorting and targeting of the proteins. Studies have shown that protein targeting in the T2SS secretion system specifically depends on the length of the hydrophobic region of the signal peptide (Driessen et al., 2001; Korotkov, et al., 2013; Luirink & Sinning, 2004).

Protein selection for targeting and secretion is determined by the signal sequence and takes place immediately following the emergence of the signal sequence from the ribosomal complex. It was suggested that in *E. coli*, the signal recognition particle (SRP) binds to the hydrophobic signal peptide and transmembrane segments. However in some cases, competition may occur between SRP and the ribosomal-associated chaperone, including SecA/SecB. Both targeting pathways converge at the translocase in the cytoplasmic compartment (Beck, et al., 2000; Driessen et al., 2001; Luirink & Sinning, 2004).

Signal Recognition Particle (SRP)

The SRP in *E. coli* is a ribonucleoprotein complex composed of Fifty-four homolog (Ffh), a GTPase, and 4.5 S RNA (Driessen et al., 2001; Herskovits, et al., 2000; Peluso et al., 2000; Zheng & Gierasch, 1997). Ffh (453 amino acid long protein) contains three domains, the N-domain (the N-terminal alpha-helical domain), the G-domain (the GTPase domain), and the M-domain (the C-terminal methionine-rich domain). The main role of the N-domain is to associate with the adjacent G-domain to create the NG domain responsible for the GTPase functionality. Although the function of Ffh is not fully understood, its function requires the presence of Mg^{2+} (Mg^{2+}-GTP) to generate a catalytically active Ffh. Assumingly, the main role of the M-domain is to connect the

SRP RNA with the nascent N-terminal signal peptide of the polypeptide (Driessen et al., 2001; Freymann, et al.r, 1997; Keenan, et al., 1998; Montoya, et al., 1997).

During co-translational translocation, the SRP recognizes and binds to the signal peptide, and creates a complex that covers the nascent N-terminal signal peptide as it emerges from the ribosome. SRP directs the complex to the SRP membrane receptor (SR), FtsY, where the conformational changes between SRP and SR complex induces the movement towards the core complex protein-conducting channel (SecYEG translocation machinery) (Collier, 1994; Low, et al., 2013; Papanikou et al., 2007). FtsY is a complex of three domains: A, N and G domains. The association of the NG domain of FtsY is related to Ffh due to their structural similarities. After the interaction between the ribosome and the SecYEG translocon occurs, the GTPase activity promotes dissociation of SRP and SR and spontaneously releases the nascent polypeptide (Driessen et al., 2001; Keenan et al., 1998; Montoya et al., 1997). This GTPase activity is based on FtsY's interaction with *E. coli* membrane phospholipids. Even though the interaction between the nascent polypeptide and the SRP is not fully understood, studies point out that the SRP RNA collaborates with SRP proteins to enable the binding and release of the signal peptide from the SRP complex (Figure 4) (Driessen et al., 2001; Freymann et al., 1997; Keenan et al., 1998; Montoya et al., 1997; Zheng & Gierasch, 1997).

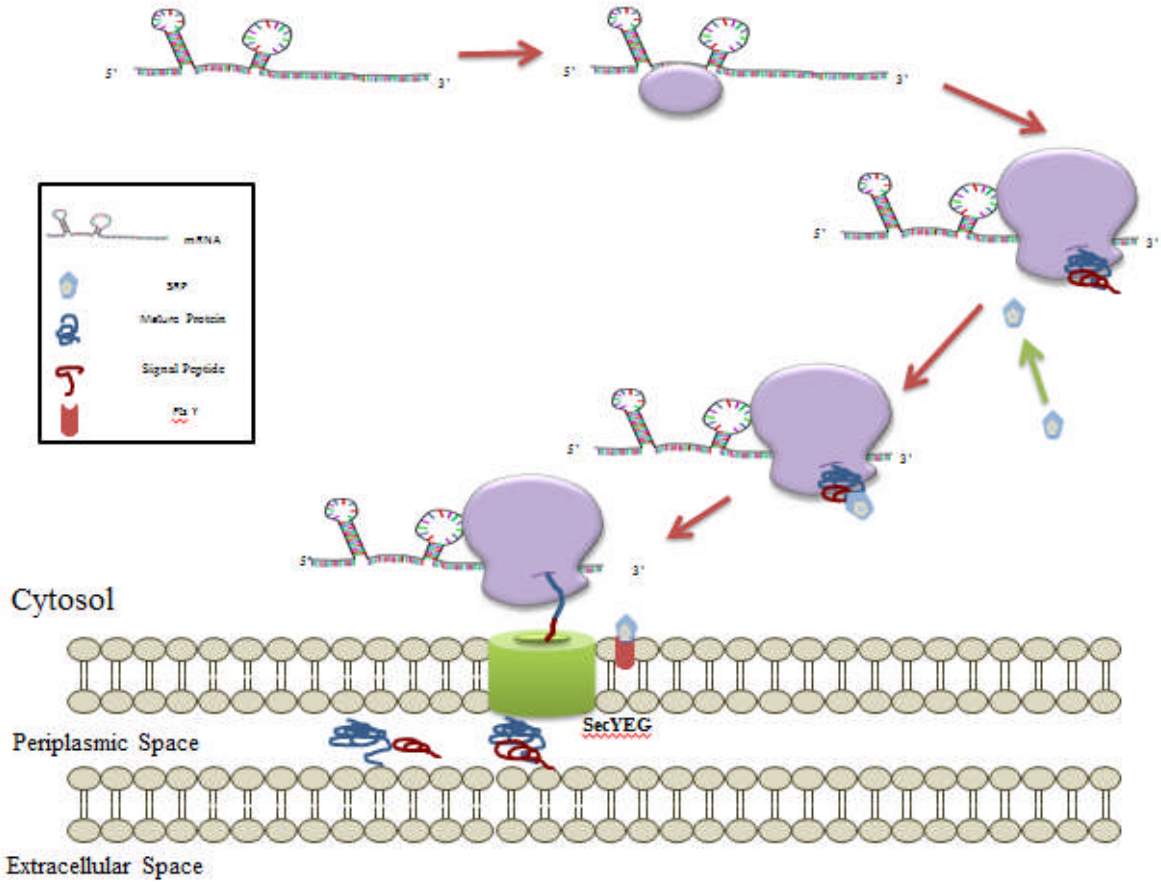

Figure 4: The current module of Sec-dependent co-translational translocation. The interaction between the SRP and the N-terminal signal peptide is during protein secretion utilizing Sec-dependant pathway (T2SS secretion system). Adopted from http://www.athenaes.com/tech_brief_ACESyebf.php

Our knowledge of the mechanistic details of the T2SS secretion system is extensive but many areas remain to be explored. For example, recent studies have shown that protein secretion efficiency may rely on signals other than the N-terminal signal sequence (Anderson & Schneewind, 1999; Samander et al., 2013). Using a bioinformatic approach, Samander et al. (2013) have shown that significant variations exist in the signal peptide sequences, a feature that may make the process of recognition and sorting of the secreted proteins challenging for the cell. Another study has shown that the T2SS secretion system is promiscuous, accepting significant variations within the sequence of the signal peptide (Samander et al., 2013). That study by Samander et al. has also shown that the equivalent nucleotides sequences within the 5′ end of the secreted protein's mRNA are more conserved and restricted in variations. This potentially allows for a higher level of efficiency in recognizing, sorting, and targeting the mRNA-ribosome complex to the SecYEG translocon. Anderson and Schneewind (1999) have reported on the role of mRNA structure in targeting the protein substrate for secretion through the T3SS secretion system.

U-richness phenomenon

Looking at the disparity of uracil distribution in various mRNAs, Prilusky and Bibi (2009) reported on the sequences that differentiate the mRNA of secretory from membrane proteins, dubbed the U-richness phenomenon. Their main observation was that the frequency of uracil varies between the specific mRNA subsets (Prilusky & Bibi, 2009). The research examined the nucleotides sequence of various secreted proteins from different organisms. The study concluded that all of the hydrophobic amino acids corresponded to a high ratio of uracil compared to other nucleotides: about 50% of the codons. A subsequent study introduced the U-richness hypothesis in which membrane proteins can be differentiated from cytoplasmic proteins by U-richness in their 5' mRNA sequence as well as pointing out the possible role of the mRNA in protein targeting

(Bibi, 2011). Similar results were reported by the study of Samander et al. where the ratio of AU/GC was observed to be higher within the 5' end of the mRNA encoding secreted protein as compared to the mRNA encoding cytoplasmic proteins.

Recognition and Sorting of mRNA

The intracellular localization of mRNA can have obvious advantages in regulating translation and targeting of proteins to specific subcellular locations (Palacios & Johnston, 2001). It is also a means of regulating protein function by restricting their location through localized translation of their mRNA (Martin & Ephrussi, 2009). Roughly up to 100 mRNAs are known to localize to different intracellular regions in eukaryotic systems. At least 25 transcripts have been shown to differentially localize in yeast cells (Palacios, 2007). During cell division in yeast, the ASH1 mRNA specifically localizes to the bud tip of daughter cells in *Saccharomyces cerevisiae* (Glisovic, et al., 2008; Martin & Ephrussi, 2009). Another example of transcript localization is that of β-actin mRNA, which is targeted to the lamellar region in fibroblasts (Martin & Ephrussi, 2009). In the growing oocyte or syncytial embryo in drosophila, mRNAs sorting has also been documented. In this case, mRNA targeting was shown to be crucial for the definition of the oocyte and the specification of embryonic axes in the developing embryo (Lasko, 2016). RNA transport has also been demonstrated in prokaryotes. In *E.coli,* a number of proteins localize to the cytoplasm, the inner membrane, and the poles of the cell through localization of their mRNA (Martin & Ephrussi, 2009; Nevo-dinur, et al., 2011). For example, labelled elF4A RNA-binding protein, bound to the *lacZ* mRNA, 5S RNA, and a short artificial untranslated RNA and was were shown to localize to different regions within the cell depending on the bound RNA (Broude, 2011; Nevo-dinur et al., 2011). The 5S RNA predominately localized to the pole or in focal regions devoid of nucleoids, while the *lacZ* mRNA was distributed in the cytoplasm. Similarly, the short

non-coding RNA was observed to localize to the cell poles (Nevo-dinur et al., 2011). With these observations, it is evident that mRNA can have specific recognition signals to mediate its sorting and targeting within the cell.

The mRNA Hypothesis

Research has shown that the alteration in the secondary structure of the signal peptide (through modification of its amino acid sequence) alters the efficiency of protein secretion but some other modifications do not affect it (Buskilay, et al., 2014). This was demonstrated in 1986 when scientists knocked out a portion from the signal peptide of a yeast enzyme, but showed that the modification did not affect its secretion (Kaiser & Botstein, 1986). Another study in 1999 also showed that mutations affected the signal peptide of *Yersinia enterocolitica* proteins that utilize the T3SS secretion system were still secreted to the same level as the wild type. These findings indicated that the amino acid sequence may not be the only signal in protein targeting and secretion (Anderson & Schneewind, 1999). A possible alternative to this sorting mechanism is that the proteins may present their mRNA secondary structure as a means of protein targeting, rather than being dependent only on the amino acid sequences of the N-terminus of the protein itself.

Rational

Survival and interaction of bacteria with their environments are dependent on their ability to secrete various proteins. For a protein to be successfully translocated to the extracellular milieu, bacteria have evolved a number of translocation systems to recognize and target proteins for secretion. Secreted proteins in both Gram positive and Gram negative bacteria are strictly dependent on the presence of the N-terminus signal peptide of secreted protein (T2SS secretion system). This system shows great promiscuity towards variations in the signal peptide sequences of most secreted proteins. Recent research into the T3SS secretion systems has recently challenged this widely accepted hypothesis and provided evidence that T3SS secretion system substrates may include mRNA secondary structure as a means of protein targeting instead relying solely on the N-terminus of the protein itself. The next step is to ask the question: is it possible that instead of the amino acid signal sequence the system utilizes the 5′ end of the mRNA (corresponding to the signal peptide in the protein) in targeting the substrate for secretion?

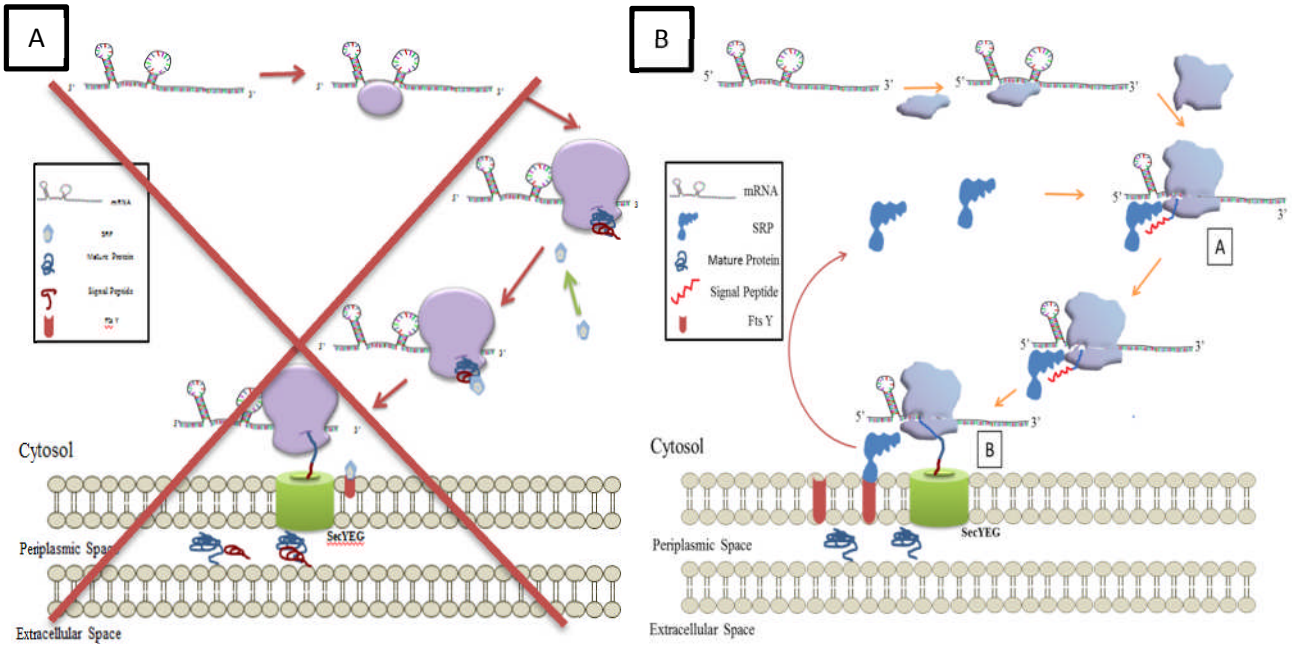

Figure 5: The current and the proposed models of Sec-dependant pathway. (A) The current model shows that the SRP binds to the N-terminal signal peptide of the nascent polypeptide emerge from the ribosomal complex and guides the complex to the translocation channel located in the membrane. (B) The proposed model indicates that the 5′ region of the mRNA secondary structure is recognized by RBP that leads the ribosomal complex to the translocation channel located in the membrane.

Hypothesis

In light of the recent research into sorting and targeting proteins for secretion, I hypothesize that the mRNA secondary structure within the 5′ end will affect the secretion efficiency of a specific secreted bacterial protein. As a test model substrate, I used the heat-stable enterotoxin b (STb) of *E. coli* and subjected its signal sequence to various mutations to test this hypothesis.

Objectives

The main research objectives of this study are:

- To clone the *E. coli* Heat-stable enterotoxin b (STb) gene because it codes for a small, secreted, and well-characterized protein.
- To introduce a number of silent and non-silent mutations in the N-terminal signal sequence of STb.
- To test whether or not these mutations affect the targeting of the toxin for secretion.

MATERIALS AND METHODS

Materials

The sequence of *stb* gene was obtained from the protein data bank (PDB) version 4 (http://www.rcsb.org/pdb/explore/explore.do?structureId=1EHS) (Figures 6). *stb* gene sequence was synthesised by Invitrogen (Thermo Fisher Scientific Inc., Toronto, Canada) and placed in a pDS32 plasmid. All bacterial culture media and chemicals for gel electrophoresis and western blots were purchased from Sigma-Aldrich Chemical Co. (Oakville, Canada).

A

5’MKKNIAFLLASMFVFSIATNAYASTQSNKKDLCEHYRQIAKESC
KKGFLGVRDGTAGACFGAQIMVAAKGC*GGGGGG**HHHHHH*-3’

B

5’**ATGAAAAAGAATATCGCATTTCTTCTTGCATCTATGTTCGTTTTT
TCTATTGCTACAAATGCCTATGCA**TCTACACAATCAAATAAAAAAG
ATCTGTGTGAACATTATAGACAAATAGCCAAGGAAAGTTGTAAAAA
AGGTTTTTTAGGGGTTAGAGATGGTACTGCTGGAGCATGCTTTGG
CGCCCAAATAATGGTTGCAGCAAAAGGATGC-3’

Figure 6: (A) STb amino acid sequence. It has been recognized by different online databases that differentiate the N-terminal signal peptide amino acid sequence from the mature secreted STb amino acid sequence. **(B) *stb* DNA sequence** “bolded and underlined” signal peptide sequence, “bolded” mature STb sequence.

mRNA Mutations

Codon Mutations (Silent Mutations (S))

The silent mutations were generated based on increasing the GC ratio in the signal peptide mRNA sequence without having any effect on the amino acids sequence. The silent mutations were applied to generate three substitution mutated sequences of the wild type *stb* signal peptide. All silent mutations were generated by site-directed mutagenesis; each “underlined” nucleotide represents the position of the nucleic acid being replaced (Table 1).

Amino Acid Mutations (Non-Silent Mutations (NS))

The NS-I mutant was designed to replace the second and the third amino acids (Lysines) with Arginines as well as increasing the GC content. The second mutation (NS-II) was designed to change the secondary structure of the mRNA as well as the hydrophobic peptide region. Thus, a hydrophobic amino acid was replaced with a charged amino acid (Leucine to Arginine). All non-silent mutations were generated by site directed mutagenesis. There are a total of two different mutated sequences of the wild type *stb* signal peptide; each “underlined” nucleotide represents the position of the nucleic acid being changed (Table 2).

Table 1: Designed (S) mutations on mRNA sequences of *stb*:

Name	mRNA Sequence
Wild type	5′AUGAAAAAGAAUAUCGCAUUUCUUCUUGCAUCUAUGUUCGUUUUUUCUAUUGCUACAAAUGCCUAUGCA 3′
Silent-I	5′AUGAAGAAGAACAUCGCGUUCCUUCUUGCAUCUAUGUUCGUUUUUUCUAUUGCUACAAAUGCCUAUGCA 3′
Silent-II	5′AUGAAGAAGAACAUCGCGUUCCUGCUGGCGUCGAUGUUCGUGUUCUCUAUUGCUACAAAUGCCUAUGCA 3′
Silent-III	5′AUGAAGAAGAACAUCGCGUUCCUGCUGGCGUCGAUGUUCGUGUUCUCGAUCGCGACGAACGCGUACGCG 3′

Table 2: Designed (NS) mutations on mRNA sequences of *stb*:

Name	mRNA Sequence
Wild type	5′AUGAAAAAGAAUAUCGCAUUUCUUCUUGCAUCUAUGUUCGUUUUUUCUAUUGCUACAAAUGCCUAUGCA 3′
Non-silent-I	5′ATGCGGCGGAAUAUCGCAUUUCUUCUUGCAUCUAUGUUCGUUUUUUCUAUUGCUACAAAUGCCUAUGCA3′
Non-silent-II	5′AUGAAAAAGAAUAUCGCAUUUCGUCUUGCAUCUAUGUUCGUUUUUUCUAUUGCUACAAAUGCCUAUGCA 3′

Bioinformatics

R-coffee

In the present project, all of the original and generated sequences were uploaded and submitted at once to the R-Coffee web-server. Multiple Sequence Alignments, MSA, is one of the formats that R-Coffee uses to arrange sequences in order to indicate the sequence homology across different sequences. Accuracy in the prototypically consistence-based methods makes MSA one of the most reliable and frequently used methods in bioinformatics and why it was chosen for this study. The sequences were uploaded in RNA format with separate names., The program is then able to recognize and align them based on the multiple methods program settings including mafft_msa, muscle_msa, and probconsRNA_msa (http://www.tcoffee.org/).

RNAfold

The RNAfold web-server was used to predict the RNA secondary structure of each sequence corresponding to each of the mutations based on the whole gene of *stb* including the signal peptide of for just the signal peptide alone. In this study, the method parameters were set to accept the minimum free energy of the structure which avoids the effect created byisolated base pairs settings and allows the program to predict only the optimal secondary structure (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

T-Coffee

The sequencing results were uploaded and submitted at once to the T-Coffee web-server in order to compare the sequence homology across different sequences. The sequences were uploaded in DNA format with separate names for each and the program was used to recognize and align them based on the original program settings with no additional multiple setting methods (http://www.tcoffee.org/).

Plasmid DNA Manipulations

Cloning

The plasmid pD444-SR, an expression vector was used in this study to express the STb protein, was purchased from [DNA2.0, USA] (Figure 8). The PCR product of the wild type *stb* was cloned into the *SnpI site* and the gene fragment was linked to the His-tag sequence by a six-glycine linker at the C-terminal of the *stb* gene. The restriction enzyme *SnpI* and T4 ligase were purchased from [New England Biolabs, USA] and utilized according to the manufacturer's recommendations. The total volume for the digestion of *stb* and pD444-SR was 15µl using *SnpI*. The ratio of vector to insert in the ligation reaction was 1:3 which means 50ng of pD444-SR and 150ng of *stb*. In this project, the molecular weight of the expressed STb protein was increased to ~6.6kDa by the addition of the 6xHIS-tag amino acid sequence and a 6 amino acid long linker at the C-terminal (Figure 9).

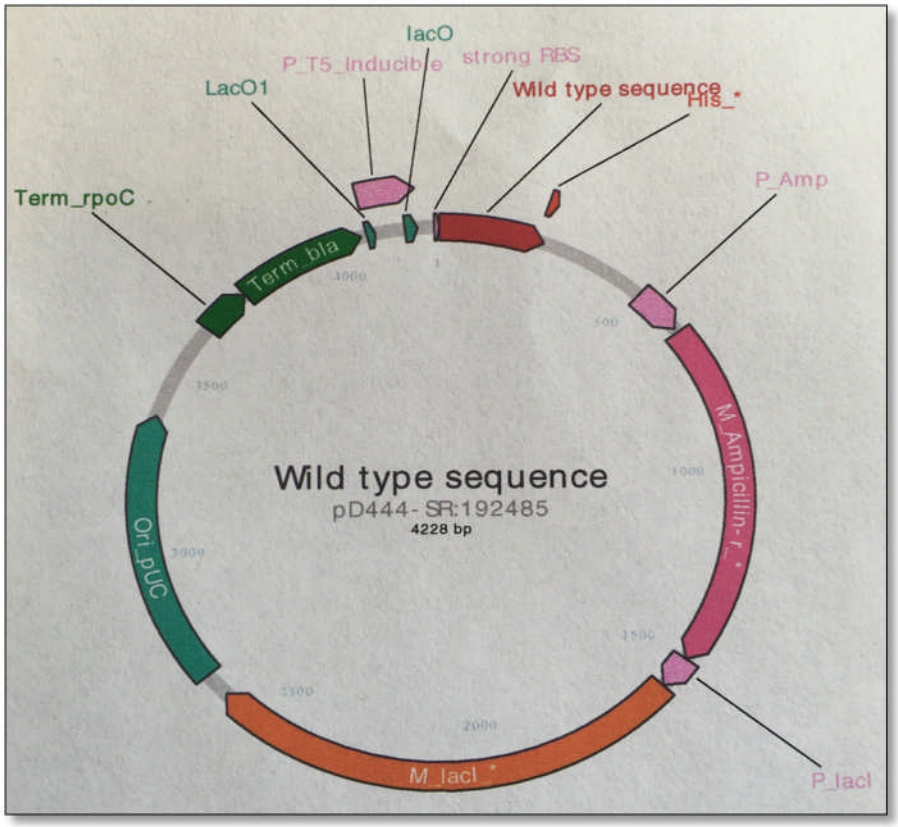

Figure 7: The Plasmid pD444-SR Vector Map including the Wild Type of *stb*.

Site Directed Mutagenesis

The Q5® Site-Directed Mutagenesis Kit was purchased from New England Biolabs [USA] and used to incorporate short mutations into the *stb* sequence according to the manufacturer's instructions. The PCR was performed using a Mastercycler Gradient thermocycler [Eppendorf, USA]. The primer pairs to incorporate the mutations were designed manually based on the Q5 high fidelity polymerase kit compatibility melting temperature and was calculated using the Ta and Tm calculator from New England Biolabs (NEB). The melting temperature calculated for the primers was between 50 and- 70°C with a 5° maximum difference between each pair. The final PCR product was 4471bp in size. The primer pairs were purchased from Invitrogen, [USA] (Table 3). In brief, the PCR was carried out with initial denaturation step for 5min at 98°C then followed by 30 cycles of: 2min at 98°C, 60 sec at an optimal annealing temperature that was calculated by -New England Biolabs calculator-, and 2min at 72°C which was followed by the final extension period of 10 min at 72°C. The optimal annealing temperature for amplification of the whole plasmid ranges between 50 and 60°C depending on the primer pair (Table 3). Each 25µl reaction contained ~50ng of plasmid DNA template of pD444-SR-*stb*, 1.25mM $MgCl_2$, 0.2mM of each dNTPs, 0.5µM forward and reverse primers, and 2 U/µl of Q5 high fidelity polymerase [New England Biolabs, USA] in each 25µl reaction. Following the PCR, the products were treated with the KLD mixture of enzymes to digest the template and ligate the product in case it was still linear. After the treatment, competent BL21 *E. coli* were transformed with each generated construct.

Bacterial Cultures

The, *E. coli* strain was cultivated in selective Luria-Broth (LB) agar plates and liquid medium [Fisher Scientific]. The LB media contains 10g of Tryptone, 5g of Yeast extract and 10g Sodium chloride per liter [Fisher Scientific]. The bacterial plates, contained LB with 15g agar added [Fisher Scientific] per liter. The medium was autoclaved at 121°C for 15 min. then placed into 55°C water bath to lower the medium temperature before adding 100µg/ml Ampicillin (Amp) antibiotic [Fisher Biotech] before pouring the selective plates. Each Petri dish contained ~25ml of media. The plates were allowed to set overnight at room temperature and then stored upside down at 4°C in a plastic bag. Cells in frozen stocks were stored at -80°C in a mixture of LB medium, 10% glycerol, and 1µg/ml of Amp. From the frozen stock, selective plates were streaked with bacteria and incubated at 37°C in a New Brunswick Scientific [USA] incubator. A single colony from the overnight plates was picked and cultured in liquid medium in an Erlenmeyer flask at 30°C in an orbital shaker [Forma Scientific, Inc.] at 250 rpm and was the source of bacterial growth, bacterial DNA, and protein extracts. Since the induced culture temperature was optimized to be 30°C, all the liquid cultures were kept at 30°C.

Transformation of *E. coli* (NEB Express (BL21)) with Constructs

E. coli were transformed with the *stb*-expressing constructs using the method described on the New England Biolab website with some modifications. The *E. coli* BL21 (DE3) strain, a protein expression host, was purchased as Chemi competent cells from ([Bio-Rad] cat. no. 1563003). The competent cells were taken from -80°C and thawed on ice for about 30 min. A mixture of about 10ng of plasmid and 200µL of competent cells equivalent to $8x10^8$ cells/ml were mixed in an ice cold 1.5ml microfuge tube by flicking and then placed on for 45 min. The tube was placed into a 42°C water bath for 90 seconds and then quickly transferred into an ice-cold water bath for 5 min. Then, 800µl of Super Optimal Broth (SOC) medium [Bio-Rad], without

antibiotics was added and the tubes incubated in a New Brunswick Scientific shaker incubator at 35°C and 250 rpm for 45 min. Cells were collected by centrifugation at 12000 xg for 5 min in an AccuSpin^{TM} Micro Centrifuge [Fisher Scientific, Germany] and resuspend in 50µl of fresh LB. The cells were then plated on two 10cm LB/Amp agar plate and then incubated at 37°C in a New Brunswick Scientific [USA] incubator for 16 hours. Clones that grew on the plates were picked and individually screened for the presence of an insert by PCR using a primer pair designed by the Primer-BLAST web-server based software to have a maximum melting temperature of 50-65°C with a maximum difference of 3°C to create a 245bp product [NCBI, 2012] [Invitrogen] (Table 3). In brief, PCR was carried out with an initial denaturing step for 15 min at 98°C followed with 40 cycles of: 1 min at 95°C, 1 min at 50°C and 2 min at 72°C, and then a final extension period of 10 min at 72°C. The pD444SR-*stb* constructs were sequenced by The Center for Applied Genomic (TCAG DNA Sequencing Facility) at Sick Kids Hospital according to their recommended protocol used for DNA sequencing.

Table 3: Primer sequences used in this study.

Primer	Sequence	Description
*stb*F1	>Forward primer 5′-GGTGGTCATATGAAAAAGAATATCGCATTTCTTC-3′	WT
*stb*F2	>Forward primer 5′-ACCACCATATGAAGAAGAACATCGCGTTCCTTCTTGCATC-3′	WT
*stb*R1	>Reverse primer 5′-TTAGTGGTGGTGATGGTGATGGCC-3′	WT
*stb*s1F	> F primer 5′-TTGCATCTATGTTCGTTTTTTC-3′	S-I
*stb*s1R	> R primer 5′-CAAGAAGGAACGCGATGTTCTTCTTCAT-3′	S-I
*stb*s2F	> F primer 5′-TCGTTTTTTCTATTGCTACAAATG-3′	S-IIa
*stb*s2R	> R primer 5′-CGAACATCGACGCCAGCAGGA-3′	S-IIa
*stb*s2Fb	> F primer 5′-ATGTTCGTGTTCTCTATTGCTAC-3′	S-IIb
*stb*s2Rb	> R primer 5′-CATCGACGCCAGCAGGAACGC-3′	S-IIb
*stb*s3F	> F primer 5′-CGCGACGAACGCGTACGCGTCTAC-3′	S-III
*stb*s3R	> R primer 5′-GCGATCGAGAACACGAACATCGACG-3′	S-III
*stb*ns1F	> F primer 5′-TTCTTCTTGCATCTATGTTCG-3′	NS-I
*stb*ns1R	> R primer 5′-GAAATGCGATATTCCGCCGCAT-3′	NS-I
*stb*ns2F	> F primer 5′ -TCGTCTTGCATCTATGTTCG-3′	NS-II
*stb*ns2R	> R primer 5′-GAAATGCGATATTCTTTTTCAT-3′	NS-II
*stb*qF	>Forward primer 5′- TCTACACAATCAAATAAAAAAG -3′	qPCR – *stb*
*stb*qR	>Reverse primer 5′- AAAGCATGCTCCAGCAGTAC -3′	qPCR – *stb*
16sqF	>Forward primer 5′- AGAGTTTGATCCTGGCTCAG -3′	qPCR – 16S rRNA
16sqR	>Reverse primer 5′- CTTGTGCGGGCCCCCGTCAATTC -3′	qPCR – 16S rRNA

Plasmid DNA Extraction

The plasmid extractions were done using the Zyppy Plasmid Minipreps Kit from [ZYMO Research] as described in the recommended protocols with few modifications. In brief, a 5ml overnight culture was centrifuged at 4500 xg for 15 min in a 15ml falcon tube in an Eppendorf 5804R Centrifuge [Backman, Germany]. The pellet was re-suspended in 1000µl of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) in a 1.5ml microfuge tube and 400µl of 7x Lysis buffer was added and mixed by inverting the tube 10 times. The mixture was neutralized by adding 800µl of cold Neutralization buffer, incubated and centrifuged at 12000 xg for 2 min in AccuSpin^{TM} Micro Centrigfuge [Fisher Scientific, Germany]. The supernatant was transferred into the provided spin-column and incubated for ~2 min then centrifuged at 6000 xg for 2 min in an AccuSpin^{TM} Micro centrifuge. The flow-through was discarded, the column washed with 500µl of Endo-Wash buffer was and then washed again with 500µl of the Wash buffer. The column was transferred to a clean 1.5 ml microfuge tube and 70µl of Elution buffer added and incubated for 5 min at 37°C and then centrifuged at 12000 xg for 2 min in an AccuSpin^{TM} Micro centrifuge and the eluted plasmid DNA collected.

Agarose Electrophoresis

The gel, composed of 1% high purity agarose [Sigma-Alrich] in 1X TAE (Tris-acetate EDTA) buffer [Prometga cat. No. V4281] containing 0.5µg/ml Ethidium bromide was cast The DNA samples were mixed with 6X loading buffer (10 mM Tris-HCl (pH 7.6) 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, and 60 mM EDTA) [Fisher Biotech] and loaded into wells formed into the gel,. The DNA was separated using electrophoresis in a QS-710 Quick Screen [IBI Scientific, USA] gel apparatus and Endura 300V power supply [Labnet International Inc., Tiwan] in TAE buffer at 80V constant for 45 min and then visualized using a Chemidoc XRS [Bio-RAD, USA].

Bacterial Cell Count

To measure bacterial number the culture was serially diluted and aliquots grown on agar plates until colonies were formed and counted. The serial dilutions were created by transferring 50µl from the 50ml culture sample to 450µl sterile water tube (labelled as 10^{-1}) and then creating 1/10 dilutions to 10^{-10}. Aliquots of the 10^{-5} to 10^{-8} dilutions were plated on agar plates and incubated, inverted at 37°C 16 - 18 hours. The Colony Forming Unites (CFU) were counted (plates that contain more than 300 colonies were excluded) and the of CFU/ml was calculated using the following formula:

$$\frac{\text{Number of Colonies}}{\text{Dilution x Volume Plated}}$$

Cell Fractionation

Culture medium fraction

Bacterial samples were inoculated into 50 ml LB in a 250 ml flask and incubated for 5 hours of induction with IPTG (Isopropyl β-D-1-thiogalactopyranoside). Each 50ml culture was centrifuged at 16000 xg for 30 min at 4°C using a model J-25I Coulter Centrifuge [Beckman, Germany] to collect the cells. The culture media was removed and filtered using a disposable vacuum 0.22µm filter system [Sarstedt]. The culture media was concentrated 10 times using an Amicon Nitrogen concentrator [Millipore, USA] with the addition of a 1000 Da NMWL ultrafiltration membrane [Millipore (76mm/dm regenerated cellulose) lot no. K5HN7268], collected in a 50ml falcon tube, and stored immediately at -80°C.

Periplasmic fraction

The amount of extraction buffer that was used was based on the pellet weight (5ml/g). The pellets were re-suspended in ice-cold sterilized sucrose buffer (0.2M Tris-HCl pH 8.0, 0.5M sterilized sucrose, and 1mM EDTA) and transferred to sterilized 1.5ml microfuge tubes. The samples were incubated on ice for 30 min and centrifuged at 14,000 xg for 10 min at 4°C using an AccuSpin^{TM} Micro centrifuge. The pellets were re-suspended in ice-cold sterile shock buffer (10mM Tris pH 8.0, 0.5mM $MgCl_2$) and incubated on ice for 30 min. The cell debris was collected by centrifugation at 13,000 xg for 10 min at 4°C using an AccuSpin^{TM} Micro centriguge. The supernatants were transferred to 1.5ml microfuge tubes. The pellets were then re-suspended in ice-cold lysis buffer (50mM Tris-HCl pH 8.0, 25mM NaCl, 5mM EDTA) and the resulting Whole cell fractions stored immediately at -80°C.

Cytoplasmic fraction

The whole cell fractions (which were subjected to the periplasmic extraction) were subjected to 5 cycles of sonication for 5 sec each using a cell disrupter [Ultrasonic, USA fitted with a microtip while on ice. Each cycle was interrupted by a one minute rest time on ice. Cell debris and membranes were removed by centrifugation at 20,000xg for 30 min at 4°C using a MRS-152 High speed refrigerated microcentrifuge [Tomy, Japan]. The supernatants were used as the source for cytoplasmic proteins.

Tricine-SDS-PAGE

Tricine Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Tricine-SDS-PAGE) gels was used to detect recombinant STb protein based on the protein size. 16% SDS-PAGE slab gels and 4% PAGE stacking gels were prepared as recommended by Schagger, 2006 protocol (Schägger, 2006): 16% separating gel (1.6ml of AB-3stock solution [48g of acrylamide [Bethesda Research Laboratories (ultrapure, lot no. 9E4204)] and 1.5g of crosslinker bisacrylamide [Fisher Biotech (lot no. 045426)] in 100ml distilled water], 1x of 3x Gel buffer (3M Tris, 0.3% SDS, pH8.9), 0.75% ammonium persulfate, 0.03% TEMED, and 0.1% Glycerol in 5ml dH2O total volume); and, 4% stacking gel (0.16%l of AB-3stock solution, 1x of 3x Gel buffer (3M Tris-HCL, 0.3% SDS, pH8.9), 0.75% ammonium persulfate, 0.03% TEMED in 2ml dH2O total volume). The separating gel was cast between two glass plates (1mm x 80mm x 70mm) overlaid with 50µl of 70% isopropanol, a left to be polymerized for ~60 min. Then the 70% isopropanol was removed and the stacking gel applied directly on the top of the separating gel with the gel comb inserted according to the manufacturer instructions [Bio-RAD, in USA]. Once the gel completely polymerized (60 min) the comb was removed and the samples were loaded into the wells. All of the STb samples were mixed with 4X Laemmli buffer at a ratio of 1:3 (buffer to sample) (277.8mM Tris pH 6.8, 4.4% SDS, 44.4% Glycerol, 0.02% Bromphenol blue, and 10% v/v β-mercaptoethanol) and boiled at 95°C for 5 min. Electrophoresis was performed in running buffer (3g Tris Base, 14.4g Glycine, and 1g SDS in 1 l) in a Bio-Rad, [USA] Mini Protean electrophoresis apparatus t at a constant 40mA and 140V for 2 hours at 4°C.

Coomassie Staining

Once the electrophoresis was complete the gels were removed from the apparatus [Bio-RAD, USA] and the separating gel incubated in ~25ml of fixation buffer (10%v/v acetic acid, 50% v/v methanol) for 30 min at room temperature on a low speed on a Roto Mix type 50800 shaker [Thermolyne, USA]. The gel was washed twice with distilled water and then incubated in ~25ml Coomassie stain buffer (10%v/v Acetic Acid, 50% v/v Methanol dH_2O, and 0.1% w/v Coomassie Briliant Blue) and kept on a Roto Mix type 50800 shaker [Thermolyne, USA] at room temperature overnight. The gels were then destained in fixation buffer for 5 hours and then in distilled water until the bands were clear. The gels were documented using the GS-800 USB densitometry [Bio-Rad, USA]

Dot Blot

Polyvinylidene fluoride (PVDF) transfer membrane 0.22µm [GElifesciences] was activated with 100% methanol and spotted with 100µl of each periplasmic sample. The PVDF membrane was cut to a desired size and defined spots were created for the sample using a pencil and ruler and then the membrane was placed on a wetted layer of extra thick blotting padswith 1x Phosphate Buffer Saline-Tween solution (PBST), (137mM NaCl, 2.7mM KCl, 10mM $Na2HPO_4$, 2mM KH_2PO_4, 0.05% Tween-20, and pH 7.4). Each sample that was applied to the membrane and the membrane left to dry. After that the membrane was reactivated using 5ml of 100% methanol for 10 min before blocking in 5% dry powdered milk in PBST for 1 hour at room temperature on a Roto Mix type 50800 shaker [Thermolyne, in USA] The membrane was washed with PBST for three times and then incubated in the diluted primary antibody, "Mouse Anti-His Antibody" [GenScript, cat. no. A00186-100], at a dilution of 1:3000 in PBSTM (PBST and 5% dry powder Milk plus 0.5% Sodium Azide) for 1 hour at room temperature.. The membrane was washed with 3 times with PBST and incubated with a dilution of 1:30000 goat anti-mouse –IgG

–alkaline phosphatase secondary antibody conjugate [Sigma, lot #SLB G 1482] in PBSTM for one hour at the room temperature. Then the membrane was washed three times with PBST and one time with Phosphate Buffer Saline (PBS) (and then developed using the the BCIP/NBT Substrate for alkaline phosphatase. It was prepared by mixing 33µl of 50mg/ml BCIP"5-bromo-4-chloro-3'-indolyphosphate" [Fisher Scientific] and 330µl of 10mg/ml NBT "nitro-blue tetrazolium" [Fisher Scientific] diluted in the 20ml AP buffer [12.1g Tris+0.12g MgCl2 in 1L of water pH 9.5]. The membrane was incubated in substrate until the purple color appeared. It was then was rinsed with water to stop the development of the background coloring.

Protein Sequencing Preparation

Tricine-SDS-PAGE gel were run and then transferred to 7x7cm PVDF transfer membrane 0.22 pore size [Millipore] previously activated by 100% methanol and washed 3 times with distilled water to remove the methanol residue. Then the membrane was immersed in 1x transfer buffer (25mMTris pH 8.3, 192mM Glycine). The proteins were transferred from theee gel to a PVDF membrane using the semi-dry transfer technique. The 16% Tricine-SDS-PAGE gel was placed on the top of PVDF transfer membrane between two layers of filter paper pads wetted in 1x transfer buffer. The proteins were transferred to the membrane at 25V/90mAmp for 30 min and the membrane immersed into PBS for further applications (N-terminal Amino Acid Sequencing).

Ni-Agarose Chromatography

Culture medium was applied to a 1ml Nickel column (1 x 5 cm) purchased from GE Lifescience [lot no. 10239469]. The recommended purification protocol was followed in order to purify His-tagged STb from the culture medium. The culture media proteins were desalted using a 60ml HiPrep 26/10 desalting column [GElifescinse, lot no. 305139] and eluted in buffer A (20mM Na_2HPO_4, 0.5M $Nacl_2$, 20mM Imidazole, pH 7.4). The 1ml Ni column was conditioned with 20ml of Buffer A. The sample was loaded onto the column at a flow rate of 1ml/min and was then washed with the Buffer A. The flow through and wash fractions were collected. The protein was eluted with buffer B (20mM Na_2HPO_4, 0.5M $Nacl_2$, 500mM Imidazole) and the fractions containing STb were collected and then applied to an Amicon spin column with 3000Da NMWL cut off [Millipore] to equilibrate the buffer with PBS (pH 7.4).

Protein Assay

The Bradford Protein Assay was used to determine the concentration of the protein in the various fractions throughout this project. The assay was performed according to the manufacturer's instructions [Bio-Basic]. In brief, 20µl of the sample was mixed with 200µl substrate for 10 min in a clear 96-well plate [Corning]. The absorbance at 595nM was determined using a FluoStar Optima plate reader [BMG Labtech, Germany] and compared to a standard curve of albumin.

N-Terminal STb sequence

The N-terminal sequence of the isolated peptides was done at Sick Kids Hospital. The protein fractions corresponding to 5 µg of STb sample from the Nickel columns were separated on 16% Tricine-SDS-PAGE gels as described above. The proteins were transferred to 0.22µm PVDF membranes using the semi-dry system for 45 min on (25V, 25W, and 190mA) as described. The targeted band was located following Ponceau S staining [Sigma] for 10 min and destaining with

distilled water. The band was cut from the membrane and sent for sequencing in a 1.5ml microfuge tube to The Center for Applied Genomic (TCAG DNA Sequencing Facility) at the Sick Kids Hospital following the protocol used by Sick Kids for the N-terminal sequencing.

STb Quantification Assay

The culture media isolated from the bacterial cultures was filtered using the disposable vacuum 0.22µm filter system [Sarstedt]. The His-Tag ELISA Detection Kit [GenScript, lot no. L00436] was used in order to quantify the amount of STb from the culture medium of different constructs as recommended by the manufacturer using the provided buffers and standards. In brief, 50µl of His Tag standards or samples were added to each well of His-tagged coated 96-well plate [Corning]. Then all of the wells were incubated with 50µl of anti-His antibody (provided) and incubated at room temperature for 2 hours. The plate was washed three times with 400µl of 1x wash solution and the plate was gently patted on a paper towel to remove any residual liquid. 100µl of the secondary antibody was added (Antibody Tracer (provided)) to each well and incubated at room temperature for 90 min. The plate was washed three times and patted dry and then 100µl of Horseradish Peroxidase colorimetric detection solution [BioShop] was added to each well. The plate was incubated in the dark until the colour appears and then 50µl stopping solution was added to each well which turns the color yellow. The absorbance was read at 450nm absorbance using FluoStar Optima plate reader [BMG Labtech, Germany].

mRNA Quantification

The rapid bacterial RNA isolation kit [Bio-Basic Inc.] was used to isolate RNA. The samples were prepared by transferring about 2×10^9 to 2×10^{10} cells from a logarithmic growth culture into 1.5 ml microfuge tubes. The sample was centrifuged at 12000 xg for 30s using

AccuSpin™ Micro centrifuge and the supernatant discarded. The cells were suspended thoroughly in 100µl of a fresh Lysozyme solution (400µg/ml in RNase-free water) and incubated at 37°C for 5 min. RNase-free filtered pipette tips were used to add 1 ml of lysis buffer (Buffer-B) and mixed gently by inverting and then incubated at room temperature for 5 min. 200µl of chloroform was added to each sample and mixed by inverting. The samples were then centrifuged at 12,000 xg for 5 min at 4°C using an AccuSpin™ Micro centrifuge and the supernatants were transferred to a new RNase-free 1.5 ml microfuge tube. 200µl of 100% ethanol was added to each sample, vortexed for 30s, and then incubated at -80°C for 45 min to increase RNA yield. The samples were centrifuged at 12,000 xg for 5 min at 4°C using an AccuSpin™ Micro centrifuge and the supernatants discarded carefully. The pellets were washed with RNase-free 75% ethanol twice by inverting for 10 times. After the samples were spun at 12,000 xg for 1 min the supernatant was discarded and the pellet air-dried for 15 min at room temperature. Finally, 30µl of RNase-free water was added to dissolve each RNA pellet which was stored at -80°C for long term storage. Then1µl of the total RNA extraction was resuspended in 499µl of RNase-free water in a 0.5ml quartz cuvette and quantified using a Shimadzu UV-2401PC Spectrophotometer [Hitachi, in Japan]. The absorbance at 260nm provided a quantitative measure of the concentration of the total RNA. The formula that was used to calculate the concentration is:

$$\{(A260) \times \text{Dilution Factor} \times 40\}$$

cDNA Template Creation

The cDNA templates were created using the High-Capacity cDNA Reverse Transcription Kit [ThermoFisher] with the addition of specific reverse primers that were designed using the Primer-BLAST web-server with melting temperatures set between 50°C and 65°C with a 3°C

maximum difference between each pair. The PCR of the cDNA was able to create a 70bp to 180bp product [NCBI, 2012], the primers were synthesized by [Invitrogen] (Table 3) as recommended by the manufacturer. The total reaction was 20µl including 10µl total master mix per Reaction (2µl from 10x RT Buffer, 0.8µl from 25x dNTP Mix (100mM), 2µl of 10µM primer, 1µl of MultiScribe Reverse Transcriptase, 4.2µl Nuclease-free H_2O) and 10µl of total RNA extraction. The cDNA reaction was carried out in Master cycler Gradient [Eppendorf, USA] with an initial 10 min at 25°C then 1 cycle for 120 min at 37°C and finally 5 min at 85°C. The reaction was then stored at -20°C.

Quantitative RT-PCR

The reactions were carried out using in the Chromo4 Real Time PCR Thermocycler [Bio-Rad, USA] and according to the recommended protocol of 2x QuantiTect SYBR Green PCR kit [Qiagen]. Primer pairs were designed using the Primer-BLAST web-server with the melting temperatures set to 50 to 65°C with a 3° maximum difference between each pair (Table 3). The total reaction was 20µL including 10µL total master mix per Reaction (10µl from 2x QuantiTect SYBR Green PCR buffer, 2µl from 10µM primer stock, 6µL Nuclease-free H2O) and 10 ul RNA. In brief, PCR was carried out with an initial denaturing step for 15 min at 98°C, followed by 40 cycles of: 1 min at 95°C, 1 min at 50°C and 2 min at 72°C, and then the final extension period of 10 min at 72°C.

RESULTS

Bioinformatics Analysis

mRNA Sequence Analysis

Multiples Sequence Alignments (MSA) by R-Coffee

The results point out the overall mismatch in nucleotide sequence between the mutated and original mRNA sequences of *stb*. The results for the designed mutations show an increase in the GC content from 32% to 57%. The mutant S-III shows the highest ratio of GC content in the N-terminal signal peptide of *stb*. S-II shows almost 50% GC content in the mRNA sequence of the N-terminal signal peptide. Interestingly, S-I and NS-I show a 1% increase in GC content between them, 39% and 40% respectively. The N-terminal signal peptide mRNA sequence of NS-II shows a 1% difference relative to the wild type signal peptide sequence (Figure 8).

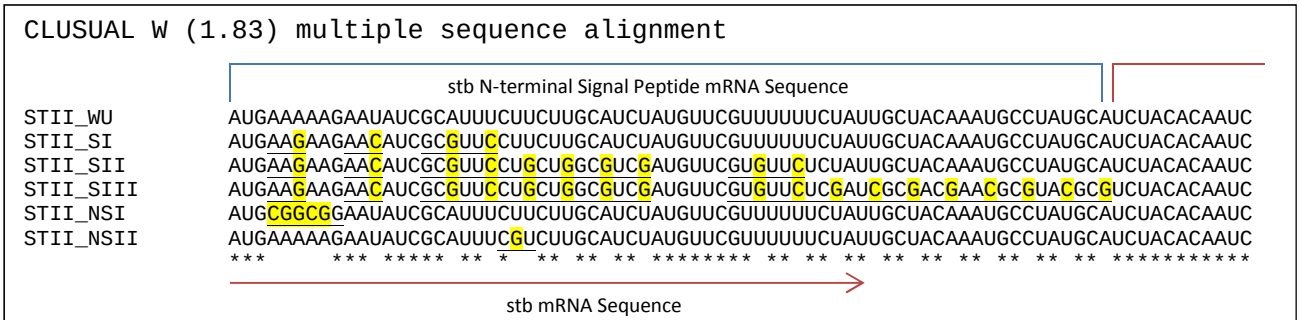

Figure 8: MSA of *stb* mRNAs by using R-Coffee web-server. The alignment of all designed mRNA sequences including the wild type. The figure clearly specifies no mismatch in the mRNA sequence of the mature STb whereas the sequence of N-terminal signal peptide shows a 23 overall varied mismatches across all mutants. The (*) in the results represents mismatched aligned nucleotides whereas No (*) means the system indicated a possible mismatch between the aligned nucleotides.

mRNA Secondary Structure Predictions

The structural simulations of all *stb* mRNA secondary structures were predicted using RNAfold web-server. The structure prediction of each *stb* mRNA sequence mutant was based on the predicted N-terminal signal peptide region. All affected and non-affected mRNAs of the *stb* signal peptide were simulated based on the minimum free energy of the computation. Therefore,

while all of the simulated sequences display an identical number of hairpins, they all show different positioning of their internal loop or loops. Collectively, the overall results of all simulations showed no identical match to either the whole mRNA sequence of the wild type *stb* or the N-terminal signal peptide mRNA sequence. All simulations show a score between (~0.4 and 1) which means the base-paring probabilities of mRNA secondary structure predictions from RNAfold web-server are reliable (Figure 9).

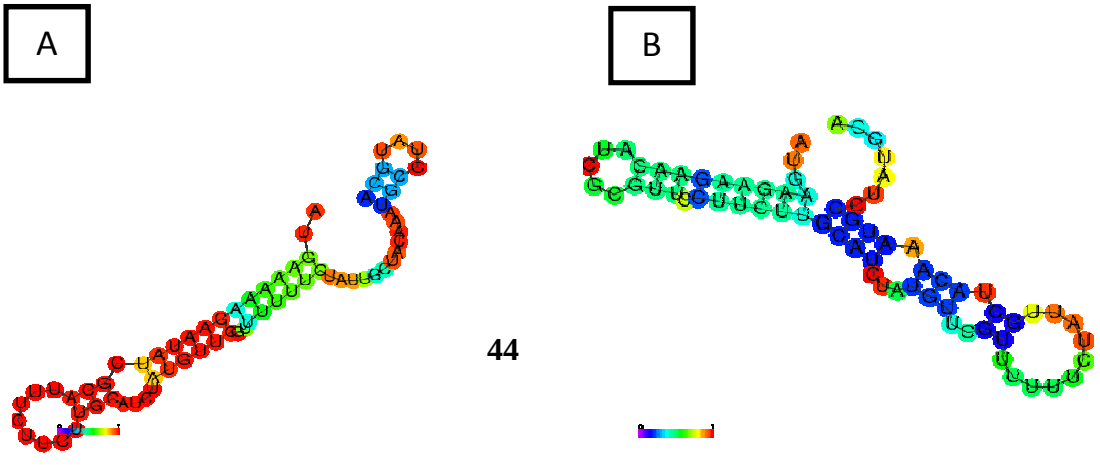

Figure 9: mRNA secondary structure prediction of signal peptide of *stb* using the RNAfold web-server. A: **Wild Type (WT), B: Silent Mutation-I (S-I), C: Silent Mutation-II (S-II), D: Silent Mutation-III (S-III), E: Non-Silent Mutation-I (NS-I), F: Non-silent mutation-II (NS-II).** Each *in silico* prediction is quite independent from the others based on different positioning of the stem loops and internal loops as well as an open reading frame. The overall coloring indicates the average confidence of base-paring probabilities.

The *stb* open reading frame was cloned directly and successfully into the pD444-SR plasmid as determined by the electrophoresis of PCR amplification products of the *stb* plasmids on 1% agarose gel which showed a band at ~240bp (Figure 10). The presence of *stb* gene into the pD444-SR plasmid was further confirmed by restriction analysis of the transformed plasmid. The

extracted construct was subjected to digestion using the *SnpI* restriction enzyme which resulted in the production of the two indicated fragments at ~300bp and at ~5000bp on 1% agarose gel electrophoresis (Figure 11) which is consistent with the sized of the vector backbone and the wildtype *stb* insert (Figure 8).

SDM was utilized to specifically generate alterations in the targeted double stranded plasmid DNA. SDM requires primers to generate the double stranded plasmid DNA: the plasmids used in these experiments are indicated in (Table 3). Plasmids extracted from the transformed E. coli BL21 bacteria indicate the presence of three different sized bands (Figure 12). The PCR results confirm the presence of the mutated *stb* in each construct that was created using the indicated primer pairs in (Table 3) (Figure 13). The control competent *E. coli* BL21 used for transformation did not show plasmid or produce bands in response to PCR amplification.

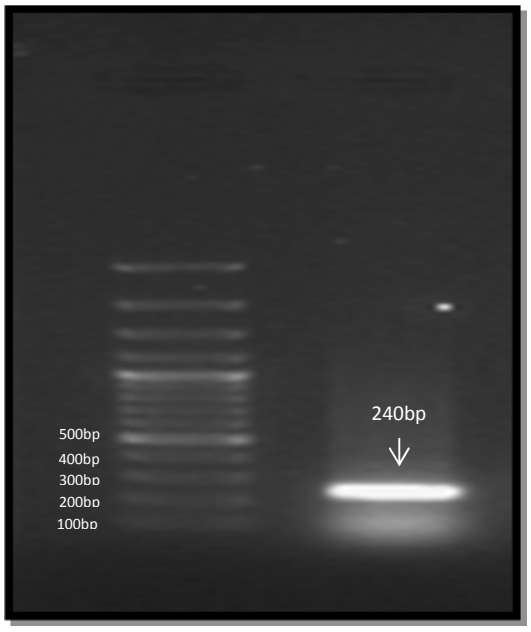

Figure 10: PCR amplification targeting *stb* gene that was resolved in 1% agarose gel shows a single expected *stb* pattern at ~240bp.

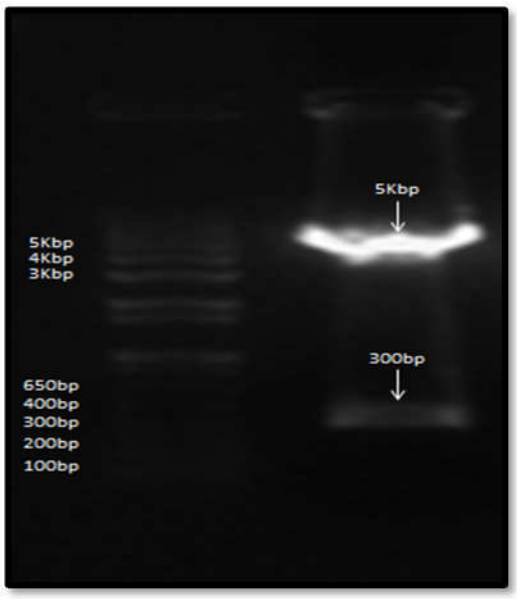

Figure 11: *stb* wild type plasmid extracted construct was subjected to a double digestion resulted into two indicated fragments on 1% agarose gel at ~300bp and at ~5000bp which are *stb* and pD444-SR respectively.

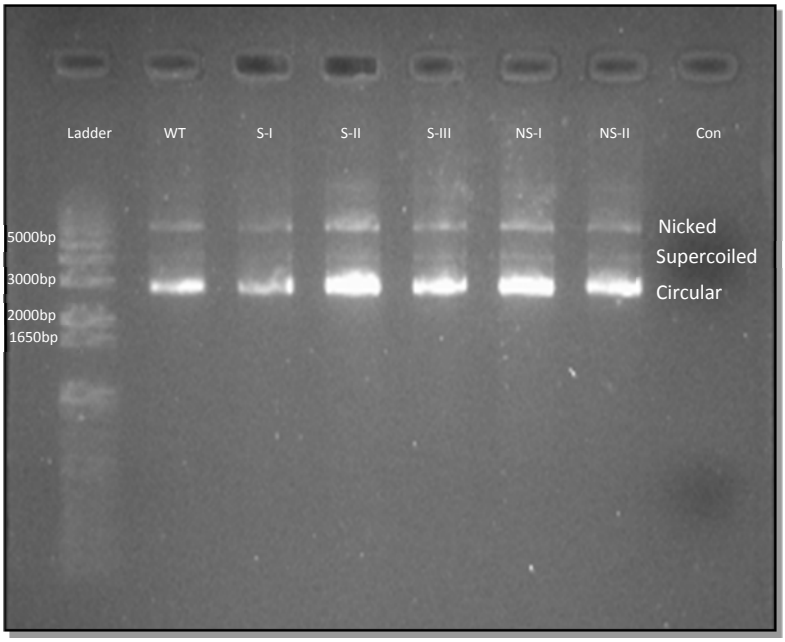

Figure 12: Plasmid extractions from all clones that were separated on 1% agarose gel indicating three different forms of plasmids (Nicked, Supercoiled and Circular) in all indicated lanes on the picture except the control lane (BL21).

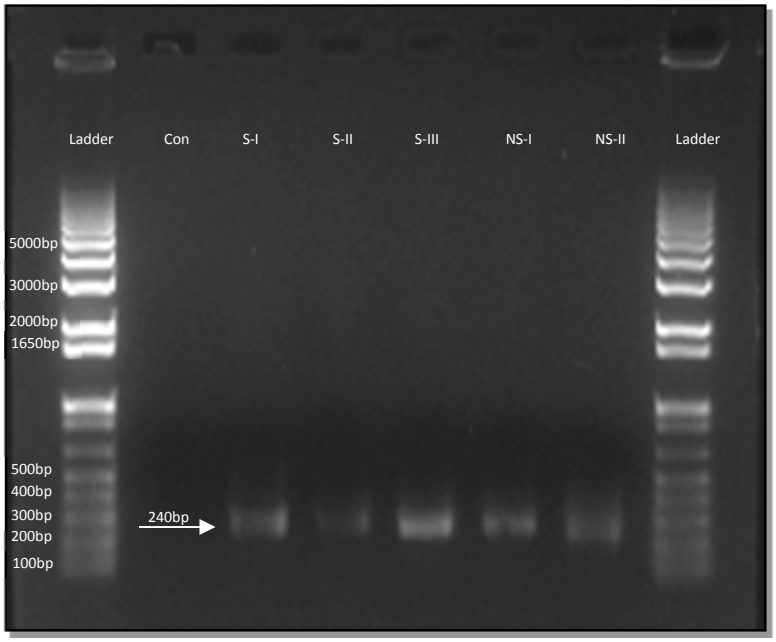

Figure 13: PCR amplifications after side-directed mutagenesis that were resolved in 1% agarose gel present a single expected *stb* pattern at ~240bp from all mutations indicated on the picture except the control *E.coli* (BL21).

Plasmid DNA Sequencing

The plasmids generated by SDM on the signal peptide of the STb were sequenced to confirm for each mutation. The samples were prepared for sequencing according to the instructions on the website for the SickKids Hospital DNA Sequencing Facility. The sequencing mixture for each sample contained 300ng of plasmid DNA and 50ng of primers and resulted in successful sequencing of the plasmids to show that the designed mutations were correct. These sequences were aligned using the T-Coffee web-server to show that the mismatched nucleotides between the mutated DNA sequences of *stb* were all located in the N-terminal signal peptide region (Figure 14).

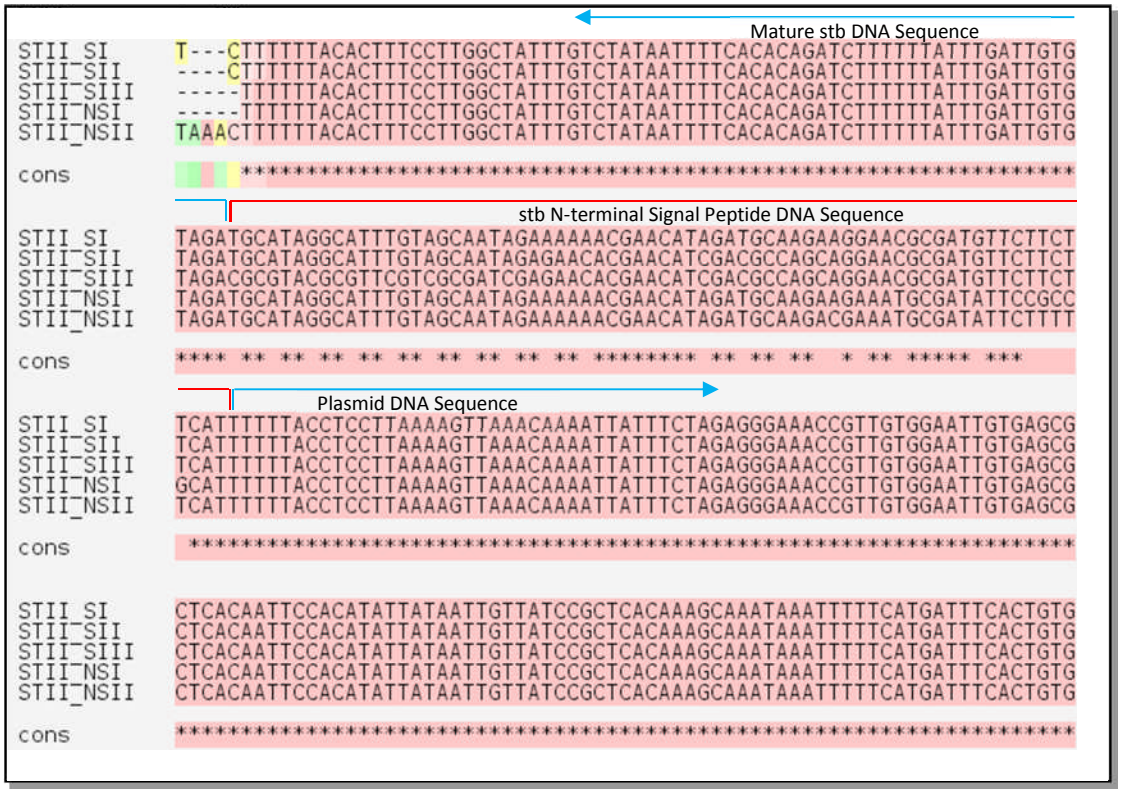

Figure 14: Alignment of Plasmid DNA sequencing results for all mutations. The sequencing results alignments of all mutations excluding the wild type were created by the Universal stb revers primer. The figure clearly indicates no mismatch in the DNA sequence of the truncated mature secreted stb as well as truncated pD444-SR plasmid sequence, whereas the signal peptide region presents an overall of 23 mismatches across all mutants as to be identical to the original design. The (*) in the results means mismatched aligned nucleotides whereas No (*) means the system indicated as a possible mismatch between the aligned nucleotides.

Protein Analysis

E. coli BL21 Growth Conditions and OD600 monitoring

Single colonies were picked for each clone from streaked LB agar/Amp plates to inoculate 5ml overnight LB medium containing 100µg/ml Amp.

The amount of bacteria from the 5 ml overnight culture needed innoculate 50 ml LB medium was measured and recorded (Table 4). The results represent the average OD_{600} of 10 overnight culture replicates which was 9.74 (+/-1.09). The optimization of the *E. coli* BL21 growth rate used in this study showed that the OD_{600} of 1.0 = 8 x 10^8 cells/ml as an average result of ten replicates. The starting culture started with ~8 x 10^7cells/ml and the OD_{600} was monitored every 60 min and recorded. An OD_{600} of 0.6 (~4.8 x 10^8cells/ml) was the point when the cultures were induced with IPTG at a specific concentration (0.4µM/ml) (Table 5). The results show that the bacteria reached the desired OD_{600} (0.6) at the same time.

STb Detection

Dot Blot

The periplasmic extraction of STb was originally done using sucrose, where the outer membrane of the bacteria stays intact (sucrose fraction), and a shock fraction where the content of the outer membrane were released and large periplasmic proteins were recovered. Confirmation for STb expression and presence were easily detected by Dot blotting the periplasmic extraction with antibodies that recognize the His epitope expressed by the recombinant protein (Figure 15).

Table 4: OD_{600} from 5 ml overnight cultures of 7 samples:

Samples	1/10 dilution	Per ml	Starting Culture	Dilution Factor	Volume needed to reach OD_{600} of 0.1
BL21	1.083	10.83	0.1	108.3	0.461ml
WT	1.144	11.44	0.1	114.4	0.437ml
S-I	1.098	10.98	0.1	109.8	0.455ml
S-II	1.086	10.86	0.1	108.6	0.460ml
S-III	1.021	10.21	0.1	102.1	0.489ml
NS-I	1.097	10.97	0.1	109.7	0.455ml
NS-II	1.053	10.53	0.1	105.3	0.474ml

Table 5: Monitoring the OD_{600} of 50ml LB culture to reach $OD_{600} \simeq 0.6$ (average of 10 replicates):

Incubation time(m)	BL21	WT	S-I	S-II	S-III	NS-I	NS-II
T=0	0.119	0.134	0.121	0.117	0.107	0.127	0.130
T=60	0.247	0.229	0.277	0.247	0.259	0.271	0.233
T=120	0.656	0.647	0.645	0.637	0.664	0.611	0.667

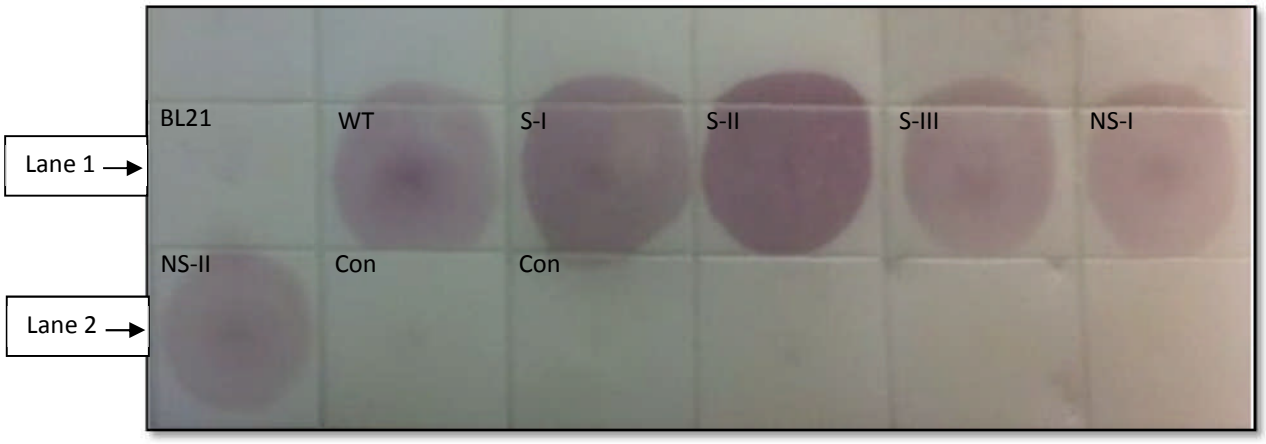

Figure 15: Dot Blot analysis of the expression of STb in *E. coli* BL21 from different clone. Dot blot analyses of detecting STb on 0.22µm PVDF membrane from periplasmic fractions are equivalent to ~200µg of total periplasmic protein. The proteins were probed with mouse anti-his primary antibody and detected with goat anti-mouse antibody conjugated with AP. Each square in lane 1 represents an indicated clone as well as lane 2, the two controls are the two extraction buffers used to extract the proteins from the periplasmic space. The results visually establish different intensity of chromogenic coloring on the PVDF membrane as a result of His-tag detection from all clones except BL21 and the controls.

Tricine-SDS-PAGE Gel

To identify STb secretion from bacterial cultures, a 10 fold concentrated sample of culture media (10µg total protein) corresponding to each clone (10 fold) was subjected to electrophoresis

on 16% Tricine-SDS-PAGE and stained with Coomassie blue stain. These gels show a similar profile of stained band except for a band of 6.6 kDa which had a remarkable different intensity between all clones. The band falls in the right range for STb and is expected to be mature secreted STb (Figure 16).

STb Identification by N-terminal Sequencing

STb was expressed, purified using the Nickel columns, and separated on 16% Tricine-SDS-PAGE gels for N-terminal sequencing. The 16% Tricine-SDS-PAGE gel showed a single band in the expected size range at approximately ~6.6 kDa (Figure 17). The gel was transferred to 0.22µm PVDF membrane using the semi-dry method and the band located following Ponceau S staining. The band was cut from the membrane and sent to the SickKids Hospital DNA sequencing facility. The results of the N-terminal protein sequencing confirmed that the band was the mature secreted STb (without the signal peptide) (Data not shown).

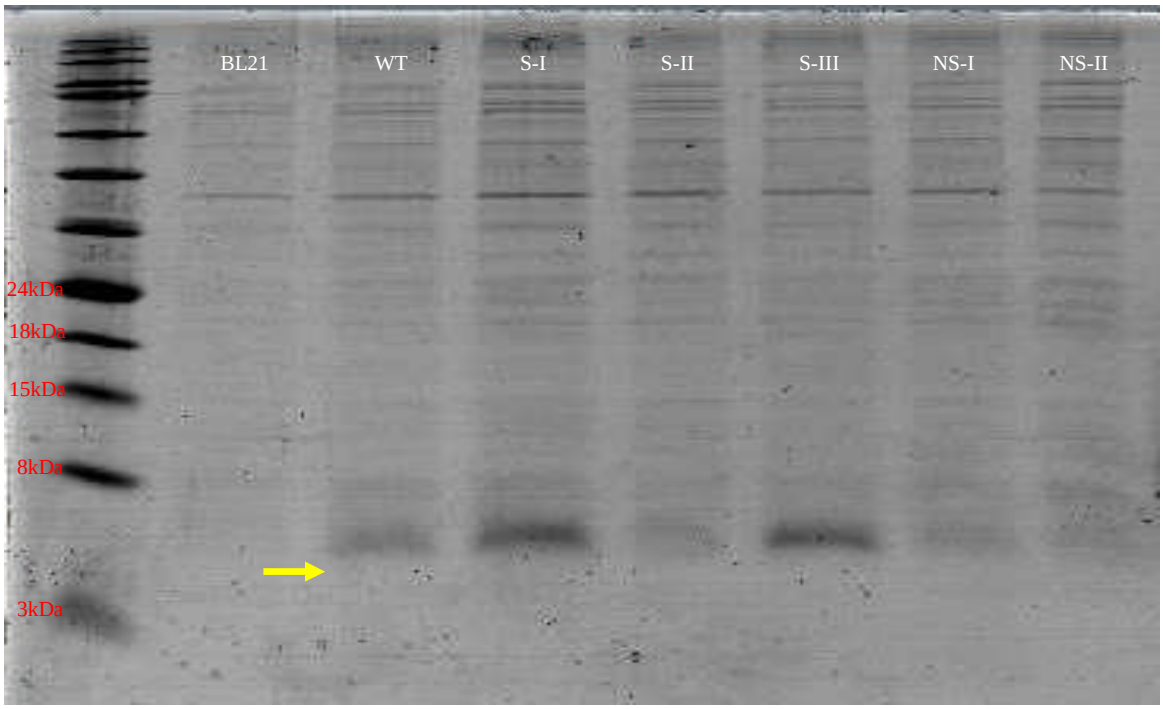

Figure 16: The expression and secretion of recombinant STb toxin from the culture media of each clone on Coomassie stained 16% Tricine-SDS-PAGE. Densitometry picture of a 10µg total secreted proteins from 50ml culture medium of each clone that is indicated on each lane on a 16% Tricine-SDS-PAGE stained by Coomassie stain. The results show that an even profile pattern intensity across all clones except one indicated ~6.6kDa band that presents a remarkable different intensity in each lane.

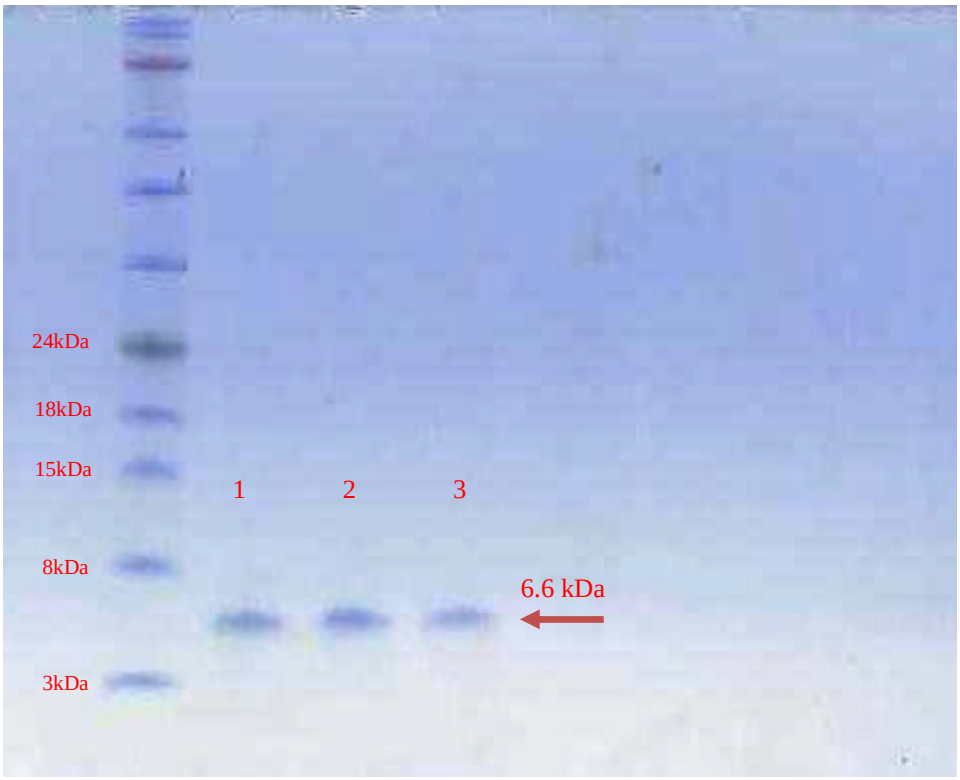

Figure 17: STb purification confirmation on Coomassie stained 16% Tricine-SDS-PAGE. The gel displays the presence of a single band in the elution fraction from the 1ml nickel column. Lane 1,2&3: Purified band (~6.6 kDa). The loading was 15µl from the elution fraction which corresponds to ~350ng of mature STb according to the N-terminal signal peptide sequencing (15µl sample + 5µl Laemmli buffer) 1:4 ratio.

STb Quantification

OD_{600} Measurement

The number of cells before induction (red color) and five hours after induction (green color) with IPTG is shown in Figure 18. The results show that the number of cells was induced to the same extent between the clones five hours after induction, except for the growth of NS-II. NS-II shows a slower growing population after induction (Figure 18). The cell count was based on *E. coli* BL21 strain which had an optimized OD_{600} of 1.0 corresponding to 8 x 10^8cells/ml.

His-tag Competitive ELISA

The presence of the 6xHis epitope tagged to the secreted STb was determined using an ELISA for the His antigen. The results were converted from an absorbance reading into quantity reading (µg/ml) based on the standard curve readings (Figure 19). The results indicate that the amount of STb present in 10µg of total proteins from the culture medium of each clone individually resulted from five triplicates subtracted from the negative control (BL21) results (Figure 20).

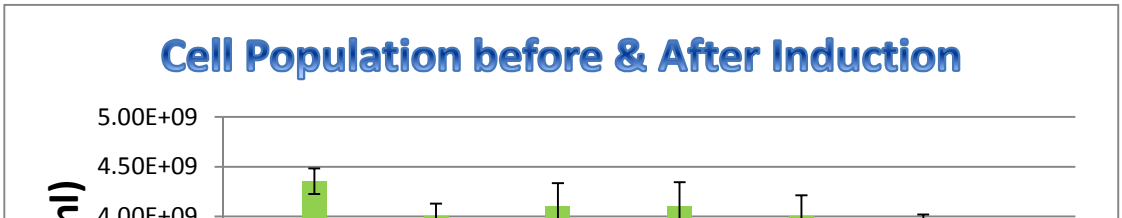

Figure 18: Cell population measurement Before & After induction by using OD_{600} method that was developed based on serial dilution cell count methods to calculate the number of the cells in each culture. The figure clearly shows an average of 5.697 x 10^8cells/ml before induction (red bars). However, all mutations show no sign of problems to achieve the same population number five hours (green bars) after induction an average of 4.063 x 10^9cells/ml except NS-II that shows reaches almost half of others cell count population in five hours an average of 2.04 x 10^9cells/ml relative to BL21. The standard error bars represents the variations of ten duplicates. The statistical results of T-test showed a significant difference with p-value of <0.001.

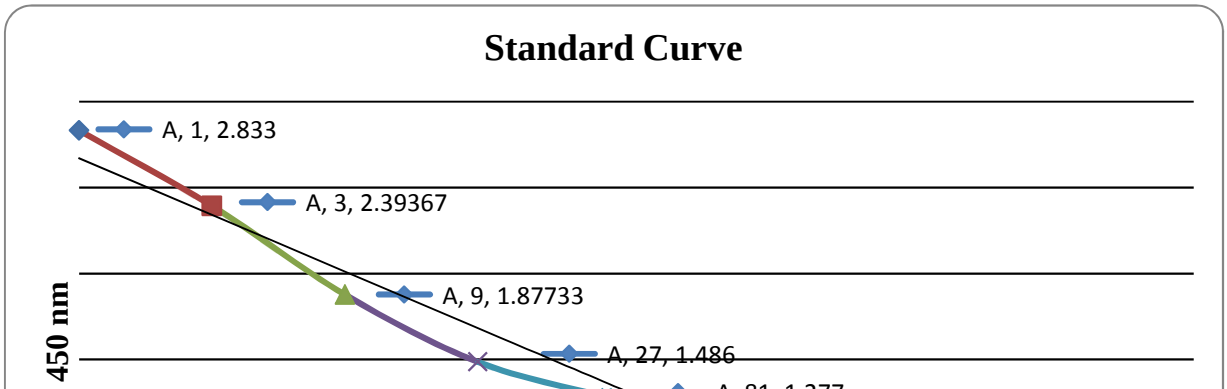

Figure 19: His-tag standard carve shows the sensitivity of the assay to detect different concentration of His-tag protein from 1ng/ml to 730ng/ml. The results show a 97% correlation of five replicates.

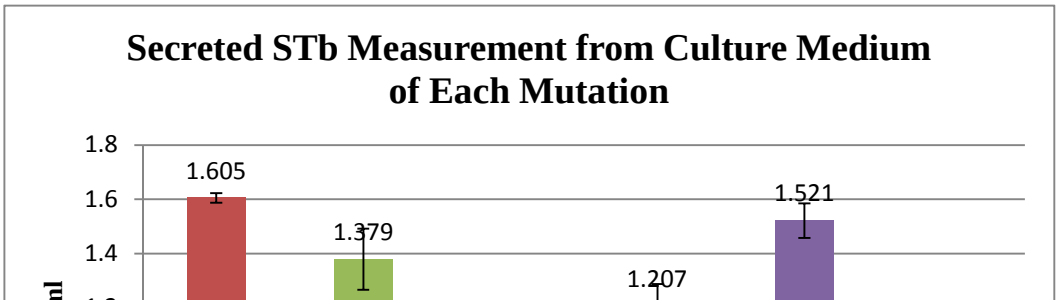

Figure 20: Competitive ELISA for secreted STb from culture medium of mutated STb clones. The results illustrate the concentration of STb from each clone that were calculated based on the His-tag standard carve. The overall results indicate an overall reduction in the presence of mature STb in the culture media relative to the WT as well as obvious variations of mature STb secretion presence between clones in 10µg total protein. The results indicate that a significant decrease in the secretion of STb from S-II and NS-II but the secretion of STb from S-I and S-III clones was slightly decreased relatively to the WT. However, NS-I clone secreted close to the level of mature WT STb. The standard error bars represents the variations of five triplicates. The statistical results of One-Way Anova showed a significant difference (*) with p-value of <0.001. The post hoc test was done using tukey test.

mRNA Quantification

qRT-PCR

The qRT-PCR was carried following successful total RNA extraction and examined for all mutant constructs compared to untransformed *E. coli* BL21. The cDNA was created using specific reverse primers for *stb* and for the 16S ribosomal protein mRNA control (Table 3). The results were normalized to the *E. coli* (BL21) as background and the 16S housekeeping mRNA to show that the expression of the stb mRNA transcript was significantly inducted by treatment with IPTG (Figure 21). The normality of the raw data was measured using Kolmogorov-Smirnov test.

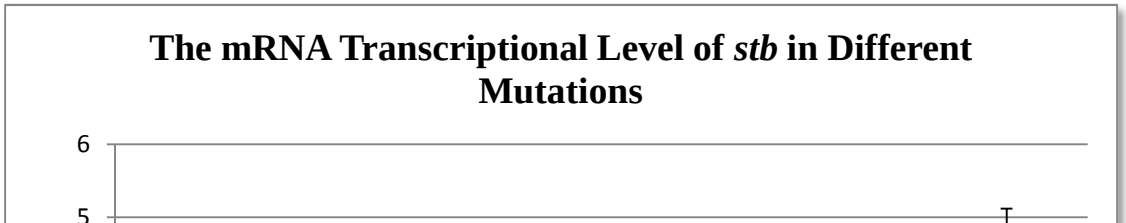

Figure 21: The relative measurement of STb mRNA transcriptional level in each clone (Different Mutation) to the wild type mRNA transcriptional level. The results indicate that the mRNA transcriptional levels of STb in all clones are relatively close to the mRNA transcriptional level of the WT STb clone (+/- 0.5 fold) except S-II and NS-II. The results S-II show that the transcriptional level of the mRNA is four times higher than the WT as well as the NS-II even higher than S-II by 1.5- 2 folds. The standard deviation bars present the relative variations of three triplicates to the WT. The results were normalized to 16S mRNA transcriptional level and subtracted from the *E. coli* (BL21). The statistical results of One-Way Anova showed a significant difference with p-value of <0.001. The post hoc test was done using tukey test.

Table 6: Summary data for the constructs. * Amount of mRNA relative to the WT. ± Represent the SD of five triplicates.

Constructs	AU:GC ratio	Amino Acid mutation	Bacterial Cell count (x10E+09)	Protein Level (µg/mL)	mRNA* Level
WT	2.1	-	4.00 ±0.2	1.6 ±0.01	1
S-I	1.6	-	4.10 ±0.5	1.4 ±0.02	1.25
S-II	1.1	-	4.10±0.5	0.1 ±0.08	3.21
S-III	0.75	-	4.00 ±0.4	1.2 ±0.08	1.01
NS-I	1.5	K2,3 →R2,3	3.83 ±0.4	1.5 ±0.06	0.58
NS-II	2.03	L8 → R8	2.04 ±0.2	0.03 ±0.1	4.49

DISCUSSION

The results of this thesis support a model of protein secretion for the STb protein that suggests that the sequence and/or secondary structure of the mRNA for Stb and the N-terminal sequence of the STb protein contribute to efficient secretion.

A variety of studies have supported the signal peptide hypothesis for secretion of proteins in E. coli. This model is widely accepted and elegantly maps how pre-secretory proteins are recognized, targeted to the Sec translocon within the plasma membrane to be processed, and eventually secreted extracellularly. For this model of protein secretion, the initial step is the recognition of the signal sequence in the translated protein by the SRP, followed by transport of the ribosome-mRNA-SRP complex to the Sec translocon. This part of the Sec pathway is referred to as the co-translational translocation of proteins. The *E. coli* STb is a small, compact protein with two disulfide bonds that are required for its activity. Early work on the mechanism of secretion indicated that it is secreted through the cotranslational translocation pathway (Forman et al., 1995; Kupersztoch et al., 1990). The secretion of STb requires the participation of SecA, DsbA (a periplasmic protein involved in disulfide formation), and TolC (an outer membrane transporter). Because of these features, STb is an ideal secretion substrate to investigate the role of the signal sequence in Sec-dependant protein secretion.

Some studies have shown experimental findings that could not be explained by the current model that describes the co-translational translocation of secreted proteins (Habyarimana & Ahmer, 2013; Samander et al., 2013). An alternate model, called the T3SS secretion system, employs an injectisome to deliver virulence factors directly to the host cell cytoplasm and utilizes a different type of a signal sequence structure than conventional T2SS model (Anderson & Schneewind, 1999). However, work on the T3SS secretion system in *Yersinia* species has recently provided evidence that the T3SS secretion system substrates may present mRNA secondary structure as a means to direct protein targeting in addition to the N-terminus of the protein itself, challenging the widely accepted hypothesis (Anderson & Schneewind, 1999). The alternate hypothesis, known as the mRNA hypothesis, proposes that the secreted polypeptides are tagged

for secretion through structural features within the 5′-end region of translated mRNA rather than through the N-terminal signal peptide signals (Habyarimana & Ahmer, 2013). Since the T2SS secretion system shows great promiscuity towards variations in the signal peptide sequences, it is possible that the T2SS system also utilizes the mRNA secondary structure in targeting the substrate (translation product) for secretion, in the same manner as the T3SS secretion system. In this work, it was hypothesized that the recognition of the 5′-end of the secreted protein's mRNA contributes to the targeting and secretion of secreted proteins.

Single-stranded RNA forms secondary structures through base pairing between complementary segments. These secondary structures may influence many cellular processes, including mRNA stability and localization, transcription, RNA processing, and translation (Katz & Burge, 2003). To propose a potential role for mRNA secondary structure in targeting and secretion of STb, it is required to show that mutations that affect secretion should also induce changes in the secondary structure of the corresponding mRNA for the secreted protein. A previous computational study proposed there was a possible specific recognition sequence that played a role in targeting secreted protein for secretion via the T2SS model. This study indicated that AU richness in the mRNA sequence within the 5′-end of the molecule played a significant role in targeting (Samander et al., 2013). In this thesis, a bioinformatics approach was used to simulate the structural consequences of mutations that were created in the signal sequence of the STb protein and in the structure of the STb mRNA. In general, the secondary structure prediction methods are based on energy minimization for double stranded structure formation from a single stranded RNA molecule. Formation of the double stranded molecule is based on nucleotide base-paring (G-C or AU). The base pairings possibilities generate different sizes of hairpins and loops within the RNA molecule.

An extensively used program for simulation of mRNA secondary structure is based on the Zuker algorithm (Zuker & Stiegler, 1981). Stability of the mRNA secondary structure is a reflection of the amount of the free energy released in forming that given structure. A program that is based on the Zuker algorithm, called RNAfold, was used in this study. Other methods used include ClustralW, MUSCLE, T-Coffee, MAFFT and PCMA (Gardner, Wilm, & Washietl, 2005). Altering the AU ratio in the signal peptide region of s*tb* mRNA indeed showed remarkable differences in loops, stem-loops and hairpins structures in most of the signal peptide mutations used in this study (Figures 9). The results showed that there is no match in the mRNA secondary structure between mutants and the wild type *stb* mRNA. However, lack of an obvious common structural feature between all the signal peptides does not support the proposition that these mRNA's would be similar enough to bind to a specific cellular factor that segregates them from each other or target them differently to specific cellular locations as compared to cytoplasmic proteins for example. The structure of the S-I mutant appears to be more of a departure from that of the wildtype sequence than the S-II mutant but the secretion level of the S-II mutant was much more reduced than secretion of S-I. Similarly, the structure of the NS-I mutant appears to be more different from the wildtype sequence but its level of secretion was closer to the wildtype level than was the NS-II mutant. This comparison is qualitative in nature and there are no methods available to reliably generate quantitative differences related to the secondary structure of these mRNAs. These simulations do not support the hypothesis that the secondary structure of the mRNA 5′-end plays a role in targeting its protein product for secretion. It is a possible point of argument however that these mRNA secondary structures were simulated using a specific algorithm and that other programs may provide different results. Other algorithms, or formation of these structures *in vivo,* may very well provide different structures than these simulations. The actual structures may, in

fact, have significant features consistent with the ability of the T2SS model to explain the ability to target and differentiate these proteins differentlyfrom each other according to the variations in secretion efficiencies for the various STb mutants used in this study (Figure 20).

To determine whether the STb had been expressed or not, dot blotting was used as the most direct indicator of STb expression. Extracting proteins from the periplasmic space is by far the fastest way to indicate the successful expression of secreted recombinant protein. The extraction protocol was adapted for its simplicity and high recovery rate for periplasmic proteins. Since STb is secreted and was tagged with a His-tag it was possible to use this technique as an initial indication of successful expression of STb in the various constructs (Figure 15). The Dot blotting results showed a gradual increase in STb secretion from WT, S-I, and S-II. However, the rest of the strains show similar levels of detectable STb to the WT. On the other hand, the controls and untransformed BL21 show no positive staining which eliminate the possibility of false positive results for the His-tag detection. However, a more quantitative method of comparison of secreted STb is required. Normalized protein concentrations from the expression of the various constructs were separated on Tricine gels. The extracts showed a distinct band at about 6.6 kDa consistent with that of the STb and the amount of the protein in this band was quantified by densitometry. For further confirmation, the band was purified by Ni-agarose and identified using N-terminal protein sequencing. The presence of the His-tag also allowed for confirmation of the results using anti-His-tag antibodies. Cell growth results show almost identical number of cells were induced in all mutations and that there was no significant difference between numbers of cells five hours after induction except for the NS-II clone. *E. coli* containing this clone seemed to grow slower after induction, taking 8 hours (instead of 5 hours) to reach the same OD_{600} as the other mutants (Figure 18). When the relative amount of STb secreted was evaluated it was clear that there was a problem

in the translation/secretion step as this clone displayed the lowest level of secretion in all the tested clones. What was unusual about this clone is that the number of copies of the mRNA was the highest in all of the tested clones. S-II shows a significant drop in the secretion level of STb in comparison to the WT and S-I while S-III showed only a minor reduction in the level of secretion. The results with the NS-II construct, in which the central hydrophobic region of the signal peptide was interrupted with a positive charge, are consistent with the current model that describes protein secretion. The essential features of the signal peptide, including the middle hydrophobic domain, are required so that this peptide can be stabilized within the membrane channel of the Sec translocon during translocation and such a disturbance in the hydrophobic region would destabilize the secondary structure of the peptide and interfere with the secretion of the protein. This clone also showed the highest levels of mRNA amongst all the clones, which is also consistent with the current model. The modification of the signal peptide structure with the addition of a positive charge within the hydrophobic domain is expected to disrupt the process of membrane insertion and translocation was rendered defective. This, in turn, caused the pausing of the translation process and the subsequent accumulation of the mRNA. These observations on the importance of the signal peptide in the process of secretion do not necessarily preclude the role of the mRNA structure in the first step in the process: that of recognition of the 5′-end of the secreted protein's mRNA in the initial step in the translation and the subsequent targeting of the ribosomal complex to the membrane, as proposed in the alternate model. Clones S-I and S-II contained a gradual increase in the GC content or a decrease in the AU ratio as compared to the wildtype sequence. What was observed was a gradual decrease in the amount of secreted STb and a gradual increase in the amount of their respective mRNAs. Again here the accumulation of the mRNAs likely reflects the reduction in the efficiency of secretion of STb in those clones. These observations are

consistent with the alternate model as proposed. The S-III clone however appears to break the rule. It has even lower AU content than the S-I and S-II constructs but the level of secretion of this mutant STb construct is higher than S-II (approaching that of the WT and S-I) and its mRNA level is also closer to that of both constructs. A plausible explanation for these results is that the sensitivity of the 5′-end of the mRNA, and its recognition for targeting to the membrane, to changes in the AU/GC ratio is restricted to a limited number of codons adjacent to the start codon of the open reading frame. The work described for the T3SS secretion model in which the mRNA structure of the 5′-end was shown to be important for secretion involved changing the AU/GC ratio in only the first 8-12 nucleotides (Anderson & Schneewind, 1999). In this study, the S-III clone had changes introduced into the AU/GC nucleotides throughout the entire signal sequence, up to nucleotide 18. Changing the AU richness causes changes to the codons and even though the amino acids sequence itself is not changed since different codons exist at different abundance levels in *E. coli*. This may impact secretion of the mutant constructs of STb used in this study. Changes in codon abundance as a result of the mutations used in this study are shown in appendix 2. In the NS-I mutant, changing the amino acids at positions 2 and 3 from K to R changes the codon abundance from 0.76 to 0.08, a very large change indeed, however STb secretion level was not affected (Figure 20). None of the other mutants constructed had changes to their translated codons as significant as this mutation and therefore it is highly unlikely that codon changes were sufficient to cause the observed effects on STb secretion seen for the S-I, S-II, or S-III mutants.

Overall, all of the designed mutations show a drop in the secretion of STb, albeit to a lesser extent for the NS-I clone. This particular clone had the least number of mutations in terms of the AU content and the essential features of the signal peptide were not modified. The most pronounced effects were observed for mutations in RNA sequence within the first 15 codons in

the 5′-end of the open reading frame. The presence of functional mutations in this region support the hypothesis that the AU/GC ratio can affect the efficiency with which a protein is secreted. This effect does not appear to be related to changes in the stability of the mRNA or to the accumulation of the toxin within the cytoplasm or the periplasmic compartment (Figure 18 & 21) but are likely due to processes specifically between the recognition of the signal sequence (in this case the mRNA is implied) and the SRP and the insertion/translocation step within the Sec translocon. The latter is unlikely to be the cause of the observed effects as the signal peptide of the nascent polypeptide is identical in sequence to that of the wild type.

CONCLUSION

This study originated from a combination of a few studies based upon the mRNA hypothesis. In this study the model I propose that the mRNA (the region corresponding to the signal sequence) secondary structure recognition is involved in the simultaneous recognition of the N-terminal signal sequence of *stb* during processing of the nascent polypeptide for secretion through the Sec-dependent T2SS. As a model substrate of the T2SS the *E. coli* STb was used to test this hypothesis. A series of mutations were created within the 5′-end of the *stb* mRNA (a region of 69 nucleotides) that corresponds to the 23 amino acid signal sequence for the T2SS. Converting

the N-terminal pair of lysines to arginines showed little effect on the growth of the bacteria or on the levels of secreted STb. A more significant mutation, however, whereby the leucine at position 8 was converted to arginine, showed a significant reduction in the amount of secreted STb. This result was anticipated as the positive charge of arginine at position 8 disrupts the hydrophobic domain of the signal peptide and potentially pausing the secretion process during the membrane insertion step within the Sec translocon. This mutant was included as a positive control and to allow us to compare the mRNA levels with the other mutations. Significantly, the sequential decrease in the AU/GC ratio within the first 15 codons showed a sequential decrease in the levels of STb secretion. The changes in the AU/GC ratios did not change the amino acid sequence of the signal peptide.

REFERENCES

Anderson, D. M., & Schneewind, O. (1999). *Yersinia enterocolitica* type III secretion: An mRNA signal that couples translation and secretion of YopQ. *Molecular Microbiology*, *31*(4), 1139–1148. http://doi.org/10.1046/j.1365-2958.1999.01254.x

Beck, K., Wu, L. F., Brunner, J., & Müller, M. (2000). Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *The EMBO Journal*, *19*(1), 134–143.

http://doi.org/10.1093/emboj/19.1.134

Beveridge, T. J. (1999). MINIREVIEW Structures of Gram-Negative Cell Walls and Their Derived Membrane Vesicles. *Journal of Bacteriology*, *181*(16), 4725–4733.

Bibi, E. (2011). Early targeting events during membrane protein biogenesis in *Escherichia coli*. *Biochimica et Biophysica Acta - Biomembranes*, *1808*(3), 841–850. http://doi.org/10.1016/j.bbamem.2010.07.025

Broude, N. E. (2011). Micro Review Analysis of RNA localization and metabolism in single live bacterial cells : achievements and challenges. *Molecular Microbiology*, *80*(5), 1137–1147. http://doi.org/10.1111/j.1365-2958.2011.07652.x

Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. M., & Wickner, W. (1990). The purified *E. coli* integral membrane protein SecYE is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell*, *62*(4), 649–657. http://doi.org/10.1016/0092-8674(90)90111-Q

Buskilay, A. A., Kannaiahy, S., & Amster-Choder, O. (2014). RNA localization in bacteria. *RNA Biology*, *11*(8), 1051–1060. http://doi.org/10.1016/j.mib.2011.01.009

Collier, D. N. (1994). *Escherichia coli* signal peptides direct inefficient secretion of an outer membrane protein (OmpA) and periplasmic proteins (maltose-binding protein, ribose-binding protein, and alkaline phosphatase) in Bacillus subtilis. *Journal of Bacteriology*, *176*(10), 3013–3020.

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and

mechanistic insights. *Nature Reviews Microbiology*, *13*(6), 343–359. http://doi.org/10.1038/nrmicro3456

de Smit, M. H., & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(19), 7668–72. http://doi.org/10.1073/pnas.87.19.7668

Driessen, A. J. M., Manting, E. H., & van der Does, C. (2001). The structural basis of protein targeting and translocation in bacteria. *Nature Structural Biology*, *8*(6), 492–498. http://doi.org/10.1038/88549

Duong, F., & Wickner, W. (1997). The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO Journal*, *16*(16), 4871–4879. http://doi.org/10.1093/emboj/16.16.4871

Franklin, R. E., & Gosling, R. G. (1953). Molecular Configuration in Sodium Thymonucleate. *Nature*. http://doi.org/10.1038/171740a0

Freymann, D. M., Keenan, R. J., Stroud, R. M., & Walter, P. (1997). Structure of the conserved GTPase domain of the signal recognition particle. *Nature*. http://doi.org/10.1038/385361a0

Gan, L., Chen, S., & Jensen, G. J. (2008). Molecular organization of Gram-negative peptidoglycan. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(48), 18953–18957. http://doi.org/10.1073/pnas.0808035105

Gardner, P. P., Wilm, A., & Washietl, S. (2005). A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Research*, *33*(8), 2433–2439.

http://doi.org/10.1093/nar/gki541

Gerlach, R. G., & Hensel, M. (2007). Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. *International Journal of Medical Microbiology*, *297*(6), 401–415. http://doi.org/10.1016/j.ijmm.2007.03.017

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Journal*, *582*(14), 1977–1986.

Greene, N. P., Porcelli, I., Buchanan, G., Hicks, M. G., Schermann, S. M., Palmer, T., & Berks, B. C. (2007). Cysteine scanning mutagenesis and disulfide mapping studies of the TatA component of the bacterial twin arginine translocase. *Journal of Biological Chemistry*, *282*(33), 23937–23945. http://doi.org/10.1074/jbc.M702972200

Habyarimana, F., & Ahmer, B. M. M. (2013). More evidence for secretion signals within the mRNA of type 3 secreted effectors. *Journal of Bacteriology*, *195*(10), 2117–2118. http://doi.org/10.1128/JB.00303-13

Harz, H., Burgdorf, K., & Höltje, J.-V. (1990). Isolation and separation of the glycan strands from murein of Escherichia coli by reversed-phase high-performance liquid chromatography. *Analytical Biochemistry*, *190*(1), 120–128. http://doi.org/10.1016/0003-2697(90)90144-X

Herskovits, A. A., Bochkareva, E. S., & Bibi, E. (2000). New prospects in studying the bacterial signal recognition particle pathway. *Molecular Microbiology*, *38*(5), 927–939. http://doi.org/10.1046/j.1365-2958.2000.02198.x

Huang, K. C., Mukhopadhyay, R., Wen, B., Gitai, Z., & Wingreen, N. S. (2008). Cell shape and

cell-wall organization in Gram-negative bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19282–7. http://doi.org/10.1073/pnas.0805309105

Ize, B., Gérard, F., Zhang, M., Chanal, A., Voulhoux, R., Palmer, T., Wu, L.-F. (2002). In vivo dissection of the Tat translocation pathway in *Escherichia coli*. *Journal of Molecular Biology*, *317*(3), 327–335. http://doi.org/10.1006/jmbi.2002.5431

Kaczanowska, M., & Rydén-Aulin, M. (2007). Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiology and Molecular Biology Reviews : MMBR*, *71*(3), 477–94. http://doi.org/10.1128/MMBR.00013-07

Kaiser, C. A., & Botstein, D. (1986). Secretion-defective mutations in the signal sequence for Saccharomyces cerevisiae invertase. *Molecular and Cellular Biology*, *6*(7), 2382–91. http://doi.org/10.1128/MCB.6.7.2382.Updated

Katz, L., & Burge, C. B. (2003). Widespread Selection for Local RNA Secondary Structure in Coding Regions of Bacterial Genes Widespread Selection for Local RNA Secondary Structure in Coding Regions of Bacterial Genes. *Genome Research*, *13*(9), 2042–2051. http://doi.org/10.1101/gr.1257503

Keenan, R. J., Freymann, D. M., Walter, P., & Stroud, R. M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell*, *94*(2), 181–191. http://doi.org/10.1016/S0092-8674(00)81418-X

Kellenberger, E., & Ryter, A. (1956). Cell Wall and Cytoplasmic Membrane of *Escherichia coli*. *Journal of Teacher Education*, *7*(3), 283–284. http://doi.org/10.1177/002248715600700323

Korotkov, K. V., Sandkvist, M., & Hol, W. G. J. (2013). The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, *10*(5), 336–351. http://doi.org/10.1038/nature08624.X-ray

Lasko, P. (2016). RNA sorting in Drosophila oocytes and embryos. *FASEB Journal*, *13*(3), 421–433.

Low, K. O., Mahadi, N. M., & Illias, R. M. (2013). Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Applied Microbiology and Biotechnology*, *97*(9), 3811–3826. http://doi.org/10.1007/s00253-013-4831-z

Luirink, J., & Sinning, I. (2004). SRP-mediated protein targeting: Structure and function revisited. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1694*(1–3 SPEC.ISS.), 17–35. http://doi.org/10.1016/j.bbamcr.2004.03.013

Lycklama a Nijeholt, J. a., & Driessen, a. J. M. (2012). The bacterial Sec-translocase: structure and mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1592), 1016–1028. http://doi.org/10.1098/r*stb*.2011.0201

Malys, N., & McCarthy, J. E. G. (2011). Translation initiation: Variations in the mechanism can be anticipated. *Cellular and Molecular Life Sciences*, *68*(6), 991–1003. http://doi.org/10.1007/s00018-010-0588-z

Marintchev, A., & Wagner, G. (2004). Translation initiation: structures, mechanisms and evolution. *Quarterly Reviews of Biophysics*, *37*(3–4), 197–284. http://doi.org/10.1017/S0033583505004026

Martin, K. C., & Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial

Dimension. *Cell*, *136*(4), 719–730. http://doi.org/10.1016/j.cell.2009.01.044

Montoya, G., Svensson, C., Luirink, J., & Sinning, I. (1997). Crystal structure of the NG domain from the signal-recognition particle receptor FtsY. *Nature*. http://doi.org/10.1038/385365a0

Natale, P., Brüser, T., & Driessen, A. J. M. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochimica et Biophysica Acta - Biomembranes*, *1778*(9), 1735–1756. http://doi.org/10.1016/j.bbamem.2007.07.015

Nevo-dinur, K., Nussbaum-shochat, A., Ben-yehuda, S., & Amster-choder, O. (2011). Translation-Independent Localization of mRNA in *E. coli*. *Science*, *331*, 1081–1085.

Nivaskumar, M., & Francetic, O. (2014). Biochimica et Biophysica Acta Type II secretion system : A magic beanstalk or a protein escalator . *BBA - Molecular Cell Research*, *1843*(8), 1568–1577. http://doi.org/10.1016/j.bbamcr.2013.12.020

Osborn, M. J. (1969). Structure and Biosynthesis of the Bacterial Cell Wall. *Digestion*, *38*, 501–538.

Palacios, I. M. (2007). How does an mRNA find its way ? Intracellular localisation of transcripts. *Seminars in Cell and Developmental Biology*, *18*, 163–170. http://doi.org/10.1016/j.semcdb.2007.01.008

Palacios, I. M., & Johnston, D. S. (2001). GETTING THE M ESSAGE A CROSS : The Intracellular Localization of mRNAs in Higher Eukaryotes. *Annual Review of Cell and Developmental Biology*, *17*, 569–614.

Palmer, T., & Berks, B. C. (2012). The twin-arginine translocation (Tat) protein export pathway.

Nature Reviews Microbiology, *10*(7), 483–496. http://doi.org/10.1038/nrmicro2814

Papanikou, E., Karamanou, S., & Economou, A. (2007). Bacterial protein secretion through the translocase nanomachine. *Nature Reviews. Microbiology*, *5*(11), 839–51. http://doi.org/10.1038/nrmicro1771

Pape, T., Wintermeyer, W., & Rodnina, M. V. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E.coli ribosome. *EMBO Journal*, *17*(24), 7490–7497. http://doi.org/10.1093/emboj/17.24.7490

Pelletier, J., & Sonenberg, N. (1987). The involvement of mRNA secondary structure in protein synthesis. *Biochemistry and Cell Biology*, *65*(6), 576–581. http://doi.org/10.1139/o87-074

Peluso, P., Herschlag, D., Nock, S., Freymann, D. M., Johnson, A. E., & Walter, P. (2000). Role of 4.5S RNA in Assembly of the Bacterial Signal Recognition Particle with Its Receptor. *Science*, *288*(5471), 1640–1643. http://doi.org/10.1126/science.288.5471.1640

Prilusky, J., & Bibi, E. (2009). Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6662–6. http://doi.org/10.1073/pnas.0902029106

Randall, L. L., & Hardy, S. J. S. (2002). SecB, one small chaperone in the complex milieu of the cell. *Cellular and Molecular Life Sciences*, *59*(10), 1617–1623. http://doi.org/10.1007/PL00012488

Rêgo, A. T., Chandran, V., & Waksman, G. (2010). Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the

chaperone-usher pathway of pilus biogenesis. *The Biochemical Journal*, *425*(3), 475–88. http://doi.org/10.1042/BJ20091518

Samander, H., Passi, K., & Saleh, M. (2013). PREDICTION OF SEC-DEPENDENT SECRETED PROTEINS BASED ON mRNA STRUCTURE OF. *BIOCOMP*, 1–7.

Sargent, F., Stanley, N. R., Berks, B. C., & Palmer, T. (1999). Sec-independent Protein Translocation in *Escherichia coli*. *Journal of Biological Chemistry*, *274*(51), 36073–36082. http://doi.org/10.1074/jbc.274.51.36073

Schägger, H. (2006). Tricine – SDS-PAGE. *Nature Protocols*, *1*(1), 16–23. http://doi.org/10.1038/nprot.2006.4

Schnaitman, C. a. (1970). Protein Composition of the Cell Wall and Cytoplasmic Membrane of *Escherichia coli*. *Journal of Bacteriology*, *104*(2), 890–901.

Thanaraj, T. A., & Pandit, M. W. (1989). An additional ribosome-binding site on mRNA of highly expressed genes and a bifunctional site on the colicin fragment of 16S rRNA from *Escherichia coli*: important determinants of the efficiency of translation-initiation. *Nucleic Acids Research*, *17*(8), 2973–85. http://doi.org/10.1093/nar/17.8.2973

Ullers, R. S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C., & Genevaux, P. (2004). SecB is a bona fide generalized chaperone in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(20), 7583–7588. http://doi.org/10.1073/pnas.0402398101

Yao, X., Jericho, M., Pink, D., & Beveridge, T. (1999). Thickness and Elasticity of Gram-Negative Murein Sacculi Measured by Atomic Force Microscopy Thickness and Elasticity

of Gram-Negative Murein Sacculi Measured by Atomic Force Microscopy. *American Society for Microbiology*, *181*(22), 6865–6875.

Yen, M. R., Tseng, Y. H., Nguyen, E. H., Wu, L. F., & Saier, M. H. (2002). Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Archives of Microbiology*, *177*(6), 441–450. http://doi.org/10.1007/s00203-002-0408-4

Zheng, N., & Gierasch, L. M. (1997). Domain interactions in *E. coli*SRP: stabilization of M domain by RNA is required for effective signal sequence modulation of NG domain. *Molecular Cell*, *1*(1), 79–87. http://doi.org/10.1016/S1097-2765(00)80009-X

Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, *9*(1), 133–148. http://doi.org/10.1093/nar/9.1.133

APPENDICES

Appendix 1

Codon usage in *E. coli*

An expanded codon table showing the relative frequency that different codons are used in *E. coli* genes is shown below.

REFERENCE: Modified from Maloy, S., V. Stewart, and R. Taylor. 1996. Genetic analysis of pathogenic bacteria. Cold Spring Harbor Laboratory Press, NY. Available at: http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/in-vitro-genetics/codon-usage.html

CODON USAGE IN *E. COLI* GENES[1]

	Codon	Amino acid[2]	%[3]	Ratio[4]	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio	
U	UUU	Phe (F)	1.9	0.51	UCU	Ser (S)	1.1	0.19	UAU	Tyr (Y)	1.6	0.53	UGU	Cys (C)	0.4	0.43	U
	UUC	Phe (F)	1.8	0.49	UCC	Ser (S)	1.0	0.17	UAC	Tyr (Y)	1.4	0.47	UGC	Cys (C)	0.6	0.57	C
	UUA	Leu (L)	1.0	0.11	UCA	Ser (S)	0.7	0.12	UAA	STOP	0.2	0.62	UGA	STOP	0.1	0.30	A
	UUG	Leu (L)	1.1	0.11	UCG	Ser (S)	0.8	0.13	UAG	STOP	0.03	0.09	UGG	Trp (W)	1.4	1.00	G
C	CUU	Leu (L)	1.0	0.10	CCU	Pro (P)	0.7	0.16	CAU	His (H)	1.2	0.52	CGU	Arg (R)	2.4	0.42	U
	CUC	Leu (L)	0.9	0.10	CCC	Pro (P)	0.4	0.10	CAC	His (H)	1.1	0.48	CGC	Arg (R)	2.2	0.37	C
	CUA	Leu (L)	0.3	0.03	CCA	Pro (P)	0.8	0.20	CAA	Gln (Q)	1.3	0.31	CGA	Arg (R)	0.3	0.05	A
	CUG	Leu (L)	5.2	0.55	CCG	Pro (P)	2.4	0.55	CAG	Gln (Q)	2.9	0.69	CGG	Arg (R)	0.5	0.08	G
A	AUU	Ile (I)	2.7	0.47	ACU	Thr (T)	1.2	0.21	AAU	Asn (N)	1.6	0.39	AGU	Ser (S)	0.7	0.13	U
	AUC	Ile (I)	2.7	0.46	ACC	Thr (T)	2.4	0.43	AAC	Asn (N)	2.6	0.61	AGC	Ser (S)	1.5	0.27	C
	AUA	Ile (I)	0.4	0.07	ACA	Thr (T)	0.1	0.30	AAA	Lys (K)	3.8	0.76	AGA	Arg (R)	0.2	0.04	A
	AUG	Met (M)	2.6	1.00	ACG	Thr (T)	1.3	0.23	AAG	Lys (K)	1.2	0.24	AGG	Arg (R)	0.2	0.03	G
G	GUU	Val (V)	2.0	0.29	GCU	Ala (A)	1.8	0.19	GAU	Asp (D)	3.3	0.59	GGU	Gly (G)	2.8	0.38	U
	GUC	Val (V)	1.4	0.20	GCC	Ala (A)	2.3	0.25	GAC	Asp (D)	2.3	0.41	GGC	Gly (G)	3.0	0.40	C
	GUA	Val (V)	1.2	0.17	GCA	Ala (A)	2.1	0.22	GAA	Glu (E)	4.4	0.70	GGA	Gly (G)	0.7	0.09	A
	GUG	Val (V)	2.4	0.34	GCG	Ala (A)	3.2	0.34	GAG	Glu (E)	1.9	0.30	GGG	Gly (G)	0.9	0.13	G
	U				C				A				G				

[1] The data shown in this table is from the Arabidopsis Research Companion on the World Wide Web (//weeds/mgh.harvard.edu). Codon frequencies for many other bacteria can be found at http://morgan.angis.su.oz.au/Angis/Tables.html.

[2] The letter in parenthesis represents the one-letter code for the amino acid.

[3] % represents the average frequency this codon is used per 100 codons.

[4] Ratio represents the abundance of that codon relative to all of the codons for that particular amino acid.

Appendix 2

Changes in codon abundance in *stb* mutant constructs:

Codon #	Abundance*					
	WT	S-I	S-II	S-III	NS-I	NS-II
1	1.00	1.00	1.00	1.00	1.00	1.00
2	0.76	0.24	0.24	0.24	0.08	0.76
3	0.24	0.24	0.24	0.24	0.08	0.24
4	0.39	0.61	0.61	0.61	0.39	0.39

5	**0.46**	**0.46**	**0.46**	**0.46**	**0.46**	**0.46**
6	**0.22**	**0.34**	**0.34**	**0.34**	**0.22**	**0.22**
7	**0.51**	**0.49**	**0.49**	**0.49**	**0.51**	**0.51**
8	**0.1**	**0.1**	**0.55**	**0.55**	**0.1**	**0.42**
9	**0.1**	**0.1**	**0.55**	**0.55**	**0.1**	**0.1**
10	**0.22**	**0.22**	**0.34**	**0.34**	**0.22**	**0.22**
11	**0.19**	**0.19**	**0.13**	**0.13**	**0.19**	**0.19**
12	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
13	**0.49**	**0.49**	**0.49**	**0.49**	**0.49**	**0.49**
14	**0.29**	**0.29**	**0.34**	**0.34**	**0.29**	**0.29**
15	**0.51**	**0.51**	**0.49**	**0.49**	**0.51**	**0.51**
16	**0.19**	**0.19**	**0.19**	**0.13**	**0.19**	**0.19**
17	**0.47**	**0.47**	**0.49**	**0.46**	**0.47**	**0.47**
18	**0.19**	**0.19**	**0.19**	**0.34**	**0.19**	**0.19**
19	**0.3**	**0.3**	**0.3**	**0.23**	**0.3**	**0.3**
20	**0.39**	**0.39**	**0.39**	**0.61**	**0.39**	**0.39**
21	**0.25**	**0.25**	**0.25**	**0.34**	**0.25**	**0.25**
22	**0.53**	**0.53**	**0.53**	**0.47**	**0.53**	**0.53**
23	**0.22**	**0.22**	**0.22**	**0.34**	**0.22**	**0.22**

*Abundance of the codon relative to all other codons for that amino acid in *E. coli*.